The frequency and associations of post endoscopy gastro-intestinal cancers

Dr Danny W F Cheung

Department of Gastroenterology, Sandwell General Hospital, West Bromwich

Thesis submitted for the high degree of Doctor of Medicine to the University of Birmingham, 2016

Institute of Cancer and Genomic Sciences

University of Birmingham

November 2016

Declaration

The entire content of this thesis is original and has not been submitted for another degree to this or any other University.

Any of the work which has been published in peer reviewed journals is acknowledged in the text or is referenced.

Acknowledgements

I wish to thank the following people for their help and support with this research, without whom this body of work would not have been possible.

I would like to thank Dr Nigel Trudgill for his supervision, patience guidance and endless encouragement for my research and drafting of the thesis.

Felicity Evison and Mr Prashant Patel at the Health Informatics department, University Hospitals Birmingham NHS Foundation Trust who have made working with Hospital Episodes Statistics possible.

I would like to thank Upper GI Blues, the local patient group at Sandwell General Hospital for their generous support in helping us to obtain the Health Improvement Network datasets used. I would also like to acknowledge the Sandwell Hospital Postgraduate Trustees for providing funds to enable the data in this body of work to be presented at national and international gastroenterological meetings.

Lastly, I would like to thank my wife Jacqueline for her continued support, without whom, this body of work would not be possible.

Table of Contents

Abbreviations

BCSP	Bowel cancer screening program
BMI	Body mass index
CRC	Colorectal cancer
EGC	Early gastric cancer
ESGE	The European Society of Gastrointestinal Endoscopy
GC	Gastric cancer
GOJ	Gastro-oesophageal junction
GORD	Gastro-oesophageal reflux disease
HES	Hospital Episodes Statistics
ICD-10	International Classification of Diseases version 10
JAG	Joint Advisory Group
KPI	Key performance indicators
NCRS	National Cancer Registration Service
NHS	National Health Service
OAC	Oesophageal adenocarcinoma
OC	Oesophageal cancer
OGD	Oesophagogastroduodenoscopy
OPCS-4	Office of Population Censuses and Surveys Classification of Interventions and Procedures (4th revision)
OSCC	Oesophageal squamous cell carcinoma
PHE	Public Health England
PCCRC	Post colonoscopy colorectal cancer
POUGIC	Post OGD upper gastrointestinal cancer

SQL	Structured Query Language
THIN	The Health Improvement Network
UGIC	Upper gastrointestinal cancer
WMCIU	West Midlands Cancer Intelligence Unit

Chapter 1: An introduction to post endoscopy gastro-intestinal cancers

1.1 Gastrointestinal cancer incidence

Gastrointestinal cancers cause significant morbidity and mortality in the UK, with the majority consisting of oesophageal cancer (OC), gastric cancer (GC) and colorectal cancer (CRC). In the UK, figures from The National Cancer Registry at the Office for National Statistics, The Information Services Division Scotland, Welsh Cancer Intelligence and Surveillance Unit and Northern Ireland Cancer Registry reported an annual incidence of over 46,000 gastrointestinal cancer cases in 2013.[1, 2, 3] Colorectal cancer (CRC) accounts for almost three quarters of all luminal gastrointestinal cancers with upper gastrointestinal cancer (UGIC), mainly consists of OC and GC accounting for the other quarter. Overall, gastrointestinal cancer accounts for every 1 in 10 new cancer diagnoses in the UK.

1.2 Oesophageal cancer

There were just under 9,000 new cases of OC registered in the UK in 2013, with male subjects being more common than female subjects in all age groups and accounting for two thirds of all OC.[1] OC is associated with older age and is relatively uncommon in patients under the age of 50 years.[4] Oesophageal adenocarcinoma (OAC) which is commonly located in the lower oesophagus and the gastro-oesophageal junction (GOJ) is the predominant histology subtype and accounts for 55% of all OC. In contrast, oesophageal squamous cell carcinoma (OSCC) is usually located at upper and middle third of the oesophagus only accounts for 28% of all OC. The incidence of OC has been on increase since

the 1970s, with an increase of 43% since the 1970s.[1, 4] The incidence of OC rose from 8.1 per 100,000 in 1998 to 10.1 per 100,000 in 2007 with the change mainly reflecting the increasing number of OAC being diagnosed with the incidence of OSCC relatively unchanged.[4] As OAC is associated with obesity, gastro-oesophageal reflux disease (GORD) and Barrett's oesophagus,[5, 6, 7] the increase in OAC incidence in the UK is in keeping with other developed nations and may be linked to increased incidence of obesity in the population.[8, 9] The overall survival rate for subjects with OC remains poor with the average 1 year survival rate at 41.9% and the 5 year survival rate of 15.1%.[1] The poor survival is likely explained by the majority of the OC subjects (73%) being diagnosed with advance disease with stage III or IV disease on diagnosis.[1]

1.3 Gastric cancer

The incidence of GC is less than OC in the UK with just over 7,000 new diagnosis recorded in 2013.[2] The incidence rate of GC is higher in male than female with over all ratio of 2:1. GC is more prevalent in older subjects with over half are diagnosed in subjects aged 75 years and age specific incidence rate shows a sharp increase around age 60 to 64 years and peaking in the 85 to 89 years age group. In contrast to OC, the GC incidence rate has declined by about two thirds since 1970s.[4] The most common GC site by anatomy recorded are overlapping or unknown site (42.4%) followed by cardia (29.0%). The overall GC survival rate is similar to that of OC with 1 year survival rate of 40% and 5 years survival rate of 20% as the majority of GC subjects also present late and 69% are diagnosed with advanced stage (III and IV) disease.[2] The survival rate is stage dependent, with the 5 years survival rate of GC subjects diagnosed with stage IA disease and stage IV disease of 80% and 5% respectively.[10]

1.4 Post OGD upper gastrointestinal cancer

Oesophagogastroduodenoscopy (OGD) is the method of choice for investigating upper gastrointestinal symptoms and for diagnosing UGIC. The image quality has improved significantly since its first development in 1960s.[11] The previous generation of fibreoptic endoscopes has been superseded by digital endoscopes and more recently by the advent of high definition digital endoscopes. However, despite the technological advancement, previous studies have reported that up to 25.8% of UGIC subjects have had an OGD which did not identify OC or GC in the 3 years prior to their eventual UGIC diagnosis.[12, 13, 14, 15, 16, 17, 18, 19] These events are termed post-OGD upper gastrointestinal cancer (POUGIC). An interval time of 3 years is used in the majority of the studies to define a false-negative endoscopic examination in the detection of UGIC. This interval is derived from the assumption that the doubling time for mucosal GC is 2 to 3 years.[20]

The overall POUGIC rates vary significantly between the published studies due to the heterogeneous nature of the POUGIC studies. The POUGIC rate is directly affected by the inclusion criteria including the time frame used to definite a POUGIC subject, the method used to calculated the POUGIC rate and also exclusion criteria such as if subjects with Barrett's oesophagus were included in the analyses. The earlier studies on POUGIC were carried out included the use of fibreoptic endoscopes which has inferior image quality and may contribute to the higher POUGIC rates reported by these studies.[17, 18, 19]

The early POUGIC studies were originated from Japan which has one of the highest incidences of UGIC in the world, with the overwhelming majority of the cases being GC.[21] Several studies by the same group assessed the diagnostic accuracy of OGD in the detection

of GC. In the first study, in which 37,094 OGD carried out between 1984 to 1989 from a single institute in Fukui Perfecture were analysed. The study reported that of the 659 GC subjects, 111 (14.4%) subjects have had previous OGD between 1 month to 3 years prior to subsequent GC diagnosis.[19] A second study by the same group analysed 9794 OGDs carried out in 1993 during which 237 subjects were diagnosed with GC on initial OGDs, with 32 (11.9%) subjects diagnosed with GC within 3 years of their initial OGD.[18] A further study compared the diagnostic accuracy between OGD and colonoscopy carried out between 1990 to 1995 but it was unclear if same subject cohort in the previous study have been included in the analysis. The study reported that GC was diagnosed initially in 542 subjects with 188 OGDs which did not detect GC initially with these subjects subsequently diagnosed with GC within 3 years. The POUGIC rate reported in this study was significantly high than other studies at 25.8%, however, this figure was calculated with the total of POUGIC OGDs rather than unique UGIC subjects therefore the POUGIC rate would be higher than comparable studies.[17]

In Western populations, early gastric cancer (EGC) is less common than in countries of the Far East and only accounts for less than 15% of all GC [15, 16, 22], where the rate of EGC may be as high as 85.2%.[19] Therefore, the POUGIC rates from the Japanese studies may not be applicable in Western populations. A number of studies from single institutes have investigated the POUGIC rate in Western populations. A Scottish study with 129 GC subjects who were referred for surgery between 1994 and 1999 reported 18 (14.0%) subjects have had OGD up to 2 years prior to their subsequent GC diagnosis.[12] The median delay for the POUGIC subjects were reported as 13 weeks. As this study only included subjects referred for a surgical opinion, the overall POUGIC rate may be higher than reported. A Finnish group

analysed OGD carried out between 1993 and 1998 with 297 GC subjects reported a POUGIC rate of 4.6% up to 3.5 years after initial negative OGD.[16] The mean delay from the first OGD to cancer diagnosis was 16.3 months in this series. A study from Oxford reviewed all OGDs carried out between 2005 and 2008 found 74 subjects were diagnosed with GC.[15] It found that 16 (21.6%) subjects have had a negative OGD up to 3 years prior to UGIC diagnosis. The higher POUGIC rate in this study may be explained by all the 10 (53%) subjects underwent OGD less than a year prior to GC diagnosis were all planned follow-ups for suspicious lesions.

Only a few studies have included both OC and GC subjects. Another study from Scotland analysed subjects undergoing OGD between 1994 to 2001 reported 30 (9.8%) subjects (16 OC and 14 GC) from a total of 305 UGIC subjects have undergone previous OGD up to 3 years prior to subsequent diagnosis.[13] Of the POUGIC subjects, 67% reportedly had a previous OGD within 1 year of UGIC diagnosis. A series from Perth, Australia reviewed OGDs carried out between 1990 to 2004 which identified a total of 822 UGIC subjects during the study period with 55 subjects having had a previous OGD. In this study, 29 of the POUGIC subjects had a negative OGD less than 1 year and 26 subjects had a negative OGD between 1 and 3 years prior to the eventual UGIC diagnosis, with a POUGIC rate (at 3 years) reported at 6.7%.[14]

The single institution studies on POUGIC were limited by the potential subject selection bias and also operator variability. There have been three studies to date which have addressed this by utilising regional and national databases. A Danish group investigated the association of anti-secretory medications (proton pump inhibitors or H2-receptor antagonists) and

POUGIC. The study analysed 513 subjects who were diagnosed with UGIC in the county of Funen between January 1993 and December 2002. It reported that 52 (10.1%) of the UGIC subjects have had a previous OGD more than 1 month prior to the subsequent UGIC diagnosis. The study did not demonstrate any significant increase risk of POUGIC between subjects who had anti-secretory medications prior to their OGD to those who were not on anti-secretory medications.[23] Two population based study from the UK by the same group analysed linked data from a national oesophao-gastric cancer audit and the national administrative health database. By using national databases, these studies addressed the major limitations from previous studies of small subject numbers from single institution with the potential patient selection bias and institutional variation.[24, 25] Both studies analysed subjects who were diagnosed with UGIC between April 2011 to March 2012 in England and Wales. They reported 537 (7.8%) of the 6943 OC subjects[24] and 225 (8.3%) of the 2727 GC subjects[25] have had a previous OGD between 3 months to 3 years prior to their UGIC diagnosis. Of the POUGIC subjects with OC, 214 (3.1%) had OGD 3-12 months and 323 (4.7%) had OGD 12-36 months prior to OC diagnosis. In subjects with GC, 102 (3.7%) and 123 (4.5%) of the POUGIC subjects have had OGD 3-12 months and 1-3 years before GC diagnosis respectively.

A recent meta-analysis which included 10 studies on POUGIC with a total of 181,662 OGD analysed, the overall UGIC prevalence was 2.1%.[26] Of the 3,787 UGIC subjects, 487 subjects had undergone an OGD which did not detect UGIC up to 3 years before diagnosis, giving an overall POUGIC rate of 12.9%. The majority of the studies in the meta-analysis only included subjects with GC, with 85.5% of all the POUGIC subjects in the meta-analysis with

GC and only 10.1% of the subjects with OC. UGIC was diagnosed approximately at 1 in every 50 OGDs with POUGIC occurring in every 400 OGDs.

1.5 POUGIC associations and risk factors

Subject demographics

In studies where subject age were analysed, younger subjects with GC were more likely to associate with POUGIC,[17, 25] however, POUGIC was not associated with age in subjects with OC.[24] There were conflicting evidence regarding gender and POUGIC with studies reporting no association[18, 24], with other studies suggesting POUGIC more likely in male subjects[14, 17] and one study reported association with female subjects.[25]

OGD indication

UGIC subjects were significantly less likely to have alarm symptoms (dysphagia, weight loss, anorexia, anaemia, haematemesis or vomiting) as the indication for OGD (32.0% to 57.2%) compared with 71.4% of controls.[13, 25] This is likely due to alarm symptoms are usually associated with advance UGIC which are more likely to be detected during endscopic examination than early malignancies. Surprisingly, one of the study reported subjects presented with alarm symptoms were two and half times more likely to associate with POUGIC.[14]

Endoscopist experience

The majority of the studies did not demonstrate any association between with occurrence of POUGIC and endoscopists experience[12, 14, 18, 19], with only one study reported

endoscopists with less than 10 years of experience were twice more likely to associate with POUGIC than more experienced endoscopists.[17]

Reported findings at POUGIC OGD

The most common findings at POUGIC OGD in GC subjects were gastritis in up to 49.8% and gastric ulcer in 14.7%.[13, 14, 25] In contrast, POUGIC subjects with OC, the most common abnormality reported on POUGIC were oesophageal ulcer in 27.3% and 18.2% with oesophageal strictures.[14] In studies where subjects with Barrett's oesophagus were included, oesophagitis or oesophageal stricture were reported in up to 56.2% in one series[13] with another study reporting oesophageal ulcer in 48.2 % of POUGIC subjects. Up to 73.1% of POUGIC subjects who had OGD within 1 years of UGIC diagnosis had an abnormality described at the site of later UGIC diagnosed was reported by an Australian series.[14] Of these subject, 15.4% had oesophageal stricture and gastric ulcers reported on their POUGIC OGD which are associated with OC and GC respectively. In POUGIC subjects with GC, gastric ulcer and gastric adenoma increase risk of subsequent GC diagnosis by five and fifty fold respectively.[18]

Biopsies sampling error

The inadequate number or lack of biopsy samples were reported as a contributing factor for POUGIC in a number of studies.[12, 13, 14, 15, 16] It was reported that up to 77% of POUGIC had inadequate biopsies (fewer than 4) taken at POUGIC OGD compared with 37% of diagnostic OGD.[13, 14] Another series also reported controls had a significantly higher median number of biopsies taken (6) when compared with POUGIC subjects (2).[15]

Previous study have reported taking between 4 to 6 biopsies provides a 97% sensitivity in detecting UGIC.[27]

UGIC location and cancer stage

In subjects with GC, there were no association between GC location and POUGIC reported by a number of studies.[17, 19, 25] However, a Japanese study reported that advance GC (T2 or greater) were more likely to be diagnosed at proximal stomach (cardia and gastric body) compare with distal stomach (angulus and antrum).[19] In contrast, OC site POUGIC subjects with OC were more likely to have OC in the proximal oesophagus compared with distal oesophagus.[24] A separate study also noted the relatively high rate of OSCC which are usually located in proximal oesophagus compared with OAC of distal oesophagus in POUGIC subjects with OC.[14] This may suggest that early OSCC maybe more difficult to detect compare with more nodular OAC and the proximal oesophagus is less well examined. Subjects with Barrett's oesophagus were more likely to be associated with POUGIC, however, endoscopic surveillance in this subject cohort is likely a major confounding factor.[24] A UK study reported that a greater proportion of POUGIC subjects with early GC (TNM stage 0 or 1)(11.5%) had OGD 3-36 months compared with POUGIC subjects with more advanced GC (TNM stage 3 or 4)(6.9%).[25] The similar finding were also reported in POUGIC subjects with OC with subjects undergoing OGD 3 to 12 months prior to their subsequent OC diagnosis more likely to have early stage disease.[24] It is impossible to know if the UGIC was present at the POUGIC OGD and overlooked by the endoscopist or if it is not detected due to a systemic failure such as inadequate biopsies samples taken or subjects not appropriately followed up.

Treatment and survival outcome of POUGIC subjects

It has been reported the POUGIC can lead to a delay in UGIC diagnosis from 0.4 years to 3.4 years.[14, 23] However, despite a few studies found a numerical survival advantage for control subjects compared with POUGIC subjects, the survival difference was not statistically significant.[15, 16, 23] This is may be due to the poor overall survival rate for UGIC subjects, with the reported average survival of less than 10 months from diagnosis in all UGIC subjects.[15, 16, 23] The poor overall survival rate is not surprising given that almost three quarters of UGIC subjects had advance disease (TNM stage 3 or 4) on diagnosis.[24, 25] A UK study did report a weak survival advantage for POUGIC subjects with OC after adjusting for cofounding factors, however, this study also included subjects with Barrett's oesophagus which may artificially improve the survival advance due to surveillance being carried out.[24] The same study also found there was no different in likelihood of undergoing treatment between POUGIC and controls after adjusting for confounding factors.

1.6 Colorectal cancer

Colorectal cancer (CRC) is far more prevalence than UGIC with over 34,000 new cases being diagnosed each year in the UK. CRC affects both genders almost equally compared UGIC with 55% CRC subjects are male.[3] The risk of CRC increases with age, with 58% of all cases diagnosed in subjects over aged 70 years and the highest incidence rates in the 85-89 years age group. CRC incidence rates have increased by 14% in the UK in the last few decades, with the recent increase likely contributed by the introduction of national bowel screening programmes over the last decade. Left sided CRC (from rectum to splenic flexure) accounts for around two thirds of all CRC.[3, 28] Significantly higher proportional of CRC subjects are diagnosed with early stage disease than subjects with OC or GC with 44% diagnosed at stage

I or II. Unsurprisingly, the survival in subjects with CRC is far superior with 1 year survival rate of 77% and 5 years survival rate of 59%.

1.7 Post colonoscopy colorectal cancer

Colonoscopy is the gold standard in investigation lower gastrointestinal symptoms and in diagnosing CRC. The standard of colonoscopy has improved dramatically since the early 2000s with quantifiable colonoscopy standards being introduced following the formation of Joint Advisory Group (JAG) on endoscopy in the UK with the quality standards being adopted in other healthcare systems since. In addition, the advent of national bowel cancer screening program (BCSP) has also improved quality of colonoscopy.[29] Measurable colonoscopy standards including caecal intubation rate and adenoma detection rate have both improved since.[30] However, studies have consistently report a proportion of CRC subjects who had colonoscopies which did not detect cancer prior to their eventual CRC diagnosis remains with these events termed post colonoscopy colorectal cancer (PCCRC). The published PCCRC rates varies according to the definition used in the studies[31], with most of the studies define PCCRC subjects having had colonoscopy examinations within 3 to 5 years prior to eventual CRC diagnosis.[28, 32, 33, 34, 35, 36]

Two of the earlier studies originated from Ontario province in Canada, using a combination of medical and insurance records and linkage to cancer registry covering the same time period from 1997 to 2001. The first study investigated 2654 subjects with right sided CRC which reported a PCCRC rate of 4.0%[32] with the subsequent study reported a PCCRC rate of 3.4% in 12,487 CRC subjects.[28] A further study from the same institute covering investigated a later period of 2000 to 2005 reported a PCCRC rate of 9.0% from 14,064 CRC

subjects.[35] Another Canadian study from Manitoba province, using hospital discharge databases and linkage to Manitoba cancer registry between 1992 to 2008 reported a PCCRC rate of 7.9% from 4883 CRC subjects.[34] A study from United States using Surveillance, Epidemiology, and End Results (SEER)-Medicare database reported PCCRC rate of 7.2% from 57839 CRC subjects during 1994 to 2005.[36]

There were two studies investigating PCCRC rate in European populations both showed significantly lower PCCRC than the North American studies. The Danish study using population-based medical registries between 2000 and 2009 reported PCCRC rate of 2.7% from 37,044 CRC subjects.[33] Similarly, a study from the UK reported PCCRC rate of 0.8% from 76,943 CRC subjects between 2006 and 2008 investigated screening interval cancer rate using a highly selected population from the national BCSP.[37]

1.8 PCCRC associations and risk factors

Subject demographics

Published PCCRC studies have reported older age[28], female gender[34], known diverticular disease [28, 36] and increased comorbidity and prior polypectomy [36] were all associated with increased risk of PCCRC.

Endoscopist factors

Studies from North American have suggested colonoscopies performed by specialty other than gastroenterology (such as an internist, general practice physician, and general surgery) were associated with high risk of PCCRC.[28, 34, 35, 36] Data also suggests endoscopist with

higher polypectomy rates were protective of later developing PCCRC when compared with endscopist with lower polypectomy rates.[35, 36] Unsurprisingly, colonoscopy carried in an office compared with in a secondary care setting were also more likely to be associated with PCCRC.[28, 35] Whether colonoscopy volume relates to PCCRC remains unclear with one study reports no association[35] with a separate study reporting increased association with higher colonoscopy volume.[36]

CRC anatomical location

Multiple studies have reported proximal location of CRC are up to twice more likely to be associated PCCRC than distal CRC.[28, 34, 35, 36]

Chapter 2: Aims and objectives

1. We hypothesised that POUGIC and PCCRC cases are associated with patient factors, endoscopist factors and provider variables.

2. To examine the incidence of POUGIC and its associations in the UK in a case-control primary care data base study.

3. To determine the incidence of POUGIC in England and also identify risk factors, provider and temporal variations using national hospital administration data.

4. To investigate the incidence of POUGIC in the Midlands and its association with patient, endoscopy and endoscopist factors using secondary care endoscopic records with cancer registry data linkage.

5. To examine the incidence of PCCRC in England and its associations in addition to provider and temporal variations using national hospital administration data.

Chapter 3: Materials and methods

3.1 The Health Improvement Network database (www.epic-uk.org)

The Health Improvement Network (THIN) database is a primary care database which includes computerised anonymised longitudinal records from over 500 primary care centres in the UK.[38] Over 5 million subjects are registered with THIN primary care centres and they are regionally and demographically representative of the UK. THIN Data collection commenced in November 2002, although retrospective data collection enables computerised records are available for research dating back to around 1988 for some patients. The data is organised by individual primary care centres and each subject is identified by a computer generated unique identifier within the centre. Participating primary care practitioners systematically record each healthcare episode as part of their routine practice, which are anonymised and prospectively recorded by the THIN software. No identifying information (such as name, address, date of birth, post code) leaves the individual primary care centre. Data are provided in four files (medical, therapy, additional health data and postcode variable indicators) with a fifth linking patient file. Clinical diagnoses are recorded in THIN as diagnostic Read codes (diagnosis dictionary) in the medical file. The therapy file contains medication records which are recorded by Multilex codes and British National Formulary coding. The additional health data file contains other health related information including height, weight, body mass index (BMI), smoking and alcohol consumptions. The post code variable file contains postcode-linked data for socio-economic status, ethnicity and environmental factors. There is a potential delay in secondary care clinical information (a new diagnosis or procedure carried out) being

recorded on the primary care system and THIN. This is reflected by an 'event date' when it occurred and a separate 'system date' when it was recorded associated with each Read code. The event date is back dated to the actual diagnosis or the procedure date. The THIN Data Collection Scheme was approved by the South-East Multicentre Research Ethics Committee (SE-MREC) as all research involving data collected from National Health Service (NHS) patients must be approved by a Research Ethics Committee.[38]

3.1.1 Study design (for Chapter 4)

A retrospective nested case-control study was performed using the THIN database to examine the incidence of POUGIC and its associations. UGIC subjects were identified as any subjects over 18 years with either a GC code (appendix 2) or OC code (appendix 3) recorded in THIN between 2002 and 2009 (for GC subjects) and 2002 and 2012 (for OC subjects). The diagnosis date of GC or OC was defined as the first record of a GC or OC diagnosis code in THIN. Cases and controls with less than 36 months of retrospective follow up available prior to their UGIC diagnosis were excluded, as it was not possible to ensure they had not undergone OGD in the 36 months prior to diagnosis. All subjects with a diagnosis of Barrett's oesophagus prior to UGIC diagnosis were also excluded to prevent confounding due to surveillance OGDs. Subjects with small intestinal cancers were not included in the study. POUGIC cases were defined as all UGIC subjects in the THIN cohort who underwent OGD between 12 and 36 months prior to eventual UGIC diagnosis. Controls were defined as UGIC subjects who did not undergo OGD between 12 to 36 months prior to UGIC diagnosis. Study variables were related to the 'diagnostic OGD' when UGIC was diagnosed in controls and the 'POUGIC OGD', the OGD which did not detect UGIC at least 1 year prior to eventual UGIC diagnosis, in POUGIC subjects. If a POUGIC subject had multiple OGDs in the 12 to 36

months period prior to their UGIC diagnosis, then the POUGIC OGD was the OGD nearest in date to when UGIC was diagnosed. In order to take into account the potential administrative delay in primary care in UGIC diagnoses being recorded in THIN, the period within 12 months of UGIC diagnosis was excluded. The POUGIC rates were calculated by dividing the number of POUGIC subjects by the total number of UGIC subjects.

3.1.2 Subjects demographics

Only birth years (rather than actual date of birth) are recorded in THIN and age was therefore rounded to the nearest whole year prior to analysis. Mean age and standard deviation was calculated to analyse the effect of age. The Charlson comorbidity index was calculated using diagnostic Read codes for medical conditions recorded in THIN prior to the diagnostic OGD date in controls and POUGIC OGD date.[39] Subjects were divided into three categories: 0 (no comorbidity), 1-4 (low comorbidity) and 5 or greater (high comorbidity). Socioeconomic status was derived at aggregate level by postcode from the subjects' place of residence. This is recorded in THIN as the Townsend deprivation index[40] and it was separated into quintiles. For the purpose of analysis, the least deprived quintiles 1 and 2 were combined and compared with quintile 3, and the most deprived quintiles 4 and 5 combined, and subjects with no recorded Townsend score. Where there was more than one Townsend score recorded in THIN, the recorded score closest to the diagnostic OGD or POUGIC OGD was used for analysis.

3.1.3 Presenting symptoms

Diagnostic Read codes for upper gastrointestinal symptoms (abdominal mass, anaemia, anorexia, dysphagia, haematemesis or melaena, gastro-oesophageal reflux disease (GORD), vomiting and weight loss) which were recorded by primary care practitioners within the 12 months prior to diagnostic OGD or POUGIC date were extracted. Alarm symptoms or signs included abdominal mass, anaemia, dysphagia, haematemesis or melaena and weight loss.

3.1.4 Endoscopic findings on POUGIC OGD

The endoscopic findings at POUGIC OGD were extracted. The endoscopic diagnoses included oesophageal stricture, oesophageal ulcer, oesophagitis, gastritis, gastric ulcer, duodenitis and duodenal ulcer. The UGIC location was recorded for OC as upper or middle oesophagus, lower oesophagus and location unknown and for GC as proximal, body, distal and location unknown. In the majority of subjects, the UGIC location was not recorded in THIN, therefore the 'free text entry' attached to the diagnostic Read code was examined to extract the anatomical location where available.

3.1.5 Treatment outcomes and survival for UGIC subjects

The number of UGIC subjects undergoing resectional surgery, chemotherapy or radiotherapy post-UGIC diagnosis were obtained by treatment Read codes. Survival was calculated from the OC or GC diagnosis date until the end of database registration, death or end of data capture in THIN, whichever was soonest. Unadjusted and adjusted (for OC or GC, gender, age, deprivation, comorbidity and alarm symptoms on presentation) survival at one year were calculated for POUGIC subjects and controls.

3.1.6 Changes in POUGIC incidence with time

In order to assess the change in the incidence of POUGIC over the study period, subjects with OC and GC were separated according to their POUGIC OGD date for cases and diagnostic date for controls into tertiles. The POUGIC rate for each tertile were then compared.

3.1.7 Statistical methodology

Statistical analysis was carried out with SPSS v20.0 (IBM, New York, USA). Independent t-test and $\chi 2$ test were used to compare differences in continuous and categorical variables respectively. Unconditional logistic regression analysis was used to calculate odds ratios and 95% confidence intervals (CI) of the influence of type of UGIC (OC or GC), gender, age, Charlson comorbidity index, socioeconomic status, presence of alarm symptoms, individual upper gastrointestinal symptoms, UGIC location, surgery, chemotherapy, radiotherapy and survival at 1 year on POUGIC. For tests of significance p values <0.05 were considered significant. A multivariate logistic regression analysis model was constructed to determine associations with POUGIC following adjusting for confounding factors including UGIC type (OC or GC), gender, age, Charlson comorbidity index, socioeconomic status and the presence of alarm symptoms. Multivariate analysis of treatment and survival outcomes were analysed by individual regression models adjusting for confounding factors including UGIC (OC or GC), gender, age, Charlson comorbidity index, socioeconomic status and the presence of alarm symptoms on presentation in each of the models. Unadjusted Kaplan-Meier analysis was used to compare survival in POUGIC subjects and controls.

3.2 Hospital Episode Statistics

Hospital Episode Statistics (HES) is an administrative database which records information on all elective and emergency episodes in NHS hospitals in England and also includes episodes of care that are NHS funded in private hospitals.[41] Since its inception in 1987, HES has expanded and currently records over 125 million episodes per year. Each care episode contains administrative information including demographic data, admission and discharge dates, admission method, diagnoses and procedures. Diagnoses are recorded using International Classification of Diseases version 10 (ICD-10), with one primary diagnosis and up to 19 secondary diagnoses coded for each episode. Procedures carried out during each episode are coded using Office of Population Censuses and Surveys Classification of Interventions and Procedures 4th revision (OPCS-4). HES is linked to Office for National Statistics (ONS) mortality records, which include date of death and all causes of death as recorded on death certificates. The ONS records therefore allow interrogation of deaths outside of hospital. The NHS provides comprehensive healthcare coverage for the UK population with the majority of OGDs and colonoscopies performed in a NHS hospital setting which are recorded in HES. [30]

3.2.1 Study design and subject definitions (for Chapter 5)

All subjects over the age of 18 years undergoing OGD between April 2003 and March 2012 were identified from HES. An OGD was defined as procedures associated with OPCS-4 codes G16 (diagnostic fibreoptic endoscopic examination of oesophagus) and G45 (diagnostic fibreoptic endoscopic examination of upper gastrointestinal tract). UGIC were identified using ICD-10 codes C15 (malignant neoplasm of oesophagus) and C16 (malignant neoplasm

of stomach) from subject diagnostic fields up to March 2015. Subjects with an UGIC diagnosis prior to the first episode of OGD were excluded from the analysis.

The date when UGIC code is recorded in HES may be delayed by a few weeks from the date of the diagnostic OGD code. For the purpose of this study, the diagnosis date was defined by the first OGD procedure code which occurred during the previous 6 months from the first cancer coding episode in HES or mortality records.[16] Subjects undergoing OGD 6 to 36 months before subsequent UGIC diagnosis were identified as POUGIC cases. These cases were further classified as POUGIC 6-12 months (OGD 6 to 12 months prior to UGIC diagnosis) and POUGIC 12-36 months (OGD 12 to 36 months prior to UGIC diagnosis). For patients who had more than one OGD between 6 to 36 months prior to UGIC diagnosis, data from the most recent OGD prior to 6 months from diagnosis the date were used for analysis, termed 'POUGIC OGD'. Controls were subjects who had not undergone OGD in the period 6 to 36 months before UGIC diagnosis, with data from the OGD when UGIC was detected, termed the 'diagnostic OGD' used for analysis. The overall POUGIC rate was calculated from the number of POUGIC subjects divided by the sum of POUGIC subjects and controls undergoing OGD between 2003 and 2012, which ensured all subjects had at least 3 years of follow up records.

3.2.2 Validation of upper gastrointestinal cancer population studied (for Chapter 5)

In order to validate the OGD population extracted from HES, to assess the validity of a diagnosis of UGIC in HES using the study methodology, the total number of UGIC cases recorded in England from the National Cancer Intelligence Network (NCIN) from 2008 to 2011 were compared with the figures from the same period coded in HES. Additionally, the

surgical and chemotherapy rates for subjects with UGIC from HES were compared with the National oesophagogastric cancer audit.

3.2.3 Study design and subject definitions (for Chapter 6)

All subjects over the age of 18 years undergoing colonoscopy including subjects in BCSP between April 2003 and March 2009 were identified from HES. Colonoscopies were defined as all procedures with OPCS-4 codes of H20.1 (snare polypectomy), H20.6 (polypectomy with colonoscopy), H22.1 (diagnostic fibreoptic endoscopic examination of colon and biopsy of lesion of colon), H22.8 (other specified diagnostic endoscopic examination of colon) and H22.9 (unspecified diagnostic endoscopic examination of colon). CRC were identified with ICD-10 codes C18 (malignant neoplasm of colon, C19 (malignant neoplasm of rectosigmoid junction) and C20 (malignant neoplasm of rectum). Subjects with a CRC diagnosis prior to the first episode of colonoscopy and subjects with a diagnosis of inflammatory bowel disease (IBD) were excluded from the analysis to avoid confounding through surveillance. Subjects with only ICD-10 code of C18.1 (malignant neoplasm of appendix) were also excluded from the analysis.

Recording of a CRC diagnosis in HES records may be delayed by a few weeks from the date of the diagnostic colonoscopy code.[42, 43] For the purpose of this study, the diagnosis date was therefore defined as the first colonoscopy code during the 6 months prior to the first CRC coding episode in HES or mortality records[35, 42], or the first CRC episode for those subjects who did not have a colonoscopy during this 6 month period due to being diagnosed through an alternative method, e.g. barium enema, CT colonography or flexible

sigmoidoscopy. Subjects undergoing colonoscopy 6 to 60 months before subsequent CRC diagnosis were identified as post colonoscopy colorectal cancer (PCCRC) cases. These cases were further classified as PCCRC 6-12 months (colonoscopy 6 to 12 months prior to CRC diagnosis); PCCRC 12-36 months (colonoscopy 12 to 36 months prior to CRC diagnosis) and PCCRC 36-60 months (colonoscopy 36 to 60 months prior to CRC diagnosis). For patients who had more than one colonoscopy 6 to 60 months prior to CRC diagnosis, data from the most recent colonoscopy was used for analysis. Controls were subjects who had not undergone colonoscopy in the period 6 to 60 months before CRC diagnosis. Colonoscopies from 2003 to 2009 were studied to ensure all subjects had at least 5 years of follow up within HES. The PCCRC rate was calculated from the number of PCCRC subjects divided by the sum of PCCRC subjects and controls.[31]

3.2.4 Validation of colonoscopy and colorectal cancer populations (for Chapter 7)

To assess the validity of the HES colonoscopy population, the number of colonoscopies between 2007 and 2010 at University Hospital Birmingham (UHB) was extracted from endoscopy records (Unisoft Medical Systems, Enfield, Middlesex, UK) and compared with the number of colonoscopies recorded in HES for UHB. To assess the validity of a CRC diagnosis in HES using the study methodology, the number of HES CRC cases was compared with the number of CRC cases diagnosed in England from the National Cancer Intelligence Network (NCIN)[44] from 2002 to 2011. Finally, the rate of surgery in the HES CRC population was compared with rate of surgery in the National Bowel Cancer Audit between 2008 and 2011.[45, 46, 47]

3.2.6 Study variables

Study variables were extracted from coding at the time of POUGIC OGD and PCCRC colonoscopy in cases. For controls, study variable were extracted from coding at diagnostic OGD for UGIC subjects and at diagnostic colonoscopy for CRC subjects.

3.2.5 Subject demographics

The recorded age of UGIC and CRC subjects was divided into tertiles according to age. Ethnicity was identified from HES demographic fields and grouped into White or White British, Asian or Asian British, Black or Black British, Chinese, Mixed and other ethnic groups.

3.2.6 Comorbidity

The Charlson comorbidity index was calculated using ICD-10 codes recorded in the diagnoses field, excluding metastatic disease as these were investigated separately, and divided into three categories: 0 (no comorbidity), 1-4 (low comorbidity) and 5 or greater (high comorbidity).[39]

3.2.7 Socio-economic status

Deprivation was assessed using the Index of Multiple Deprivations 2007, which is calculated from an aggregate score for each English catchment area derived from seven domain scores (income, employment, health deprivation and disability, education, skills and training, barriers to housing and services, crime and living environment). Individual subjects were linked to their corresponding catchment area and associations with deprivation were then analysed in quintiles, with quintile 1 being the most deprived.

3.2.8 Upper gastrointestinal cancer variables

UGIC sites were classified into proximal oesophagus, distal oesophagus, oesophagus unspecified, proximal stomach, mid stomach, distal stomach and stomach unspecified according to ICD-10 codes (appendix 4). Initial unspecified site UGIC diagnoses were interrogated to determine whether a more specific code had been used subsequently, and where this was the case, the later more specific code was used to determine the UGIC site.

ICD10 codes were used to identify oesophageal ulcers (K22.1 Ulcer of oesophagus), oesophageal strictures (K22.2 Oesophageal obstruction) in subjects with OC and gastric ulcers (K25 Gastric ulcer; K27 Peptic ulcer, site unspecified; K28 Gastrojejunal ulcer) in subjects with GC from the POUGIC OGD which were then compared with equivalent frequencies coded in all OGDs within HES.

3.2.9 Colorectal cancer variables

CRC site was classified based on the first CRC coding episode into: right sided, left sided and unspecified (appendix 5). Coding records of initially unspecified site CRC were examined and if a more specific code had been used subsequently, this was used to determine the CRC site. Colonic polyps were identified from ICD-10 codes (appendix 6).

3.2.10 Distant metastases

Distant metastases were identified by ICD-10 codes (appendix 7) up to 12 months from diagnosis date and were used as a surrogate marker of UGIC and CRC stage at diagnosis, as TNM and Dukes' staging is not recorded in HES. Codes for metastases can occasionally be

miscoded as a primary neoplasm (e.g. lung), and therefore primary malignancy codes were also used, provided that they were recorded in the 12 months subsequent to UGIC or CRC diagnosis.

3.2.11 Treatment and survival outcome

Surgery, chemotherapy and radiotherapy were identified by respective OPCS-4 codes (appendices 7, 8 and 9). Survival analysis adjusted for gender, age, deprivation and comorbidity was calculated from the diagnosis date for POUGIC cases, PCCRC cases and controls. Subjects who were not diagnosed by OGD or colonoscopy were not included in the survival analysis to avoid potential lead time bias due to the method of determining date of diagnosis from HES.

3.2.12 Provider variables for POUGIC study (Chapter 5)

All endoscopy units operating within an NHS Hospital trust were analysed as a single provider for the purpose of this chapter. Individual providers were stratified by OGD volume, the percentage of UGIC patients diagnosed by an emergency procedure rather than electively, the percentage of patients who underwent a follow up OGD following a diagnosis of gastric ulcer within 12 weeks and the provider PCCRC rate to determine if there was an association with POUGIC. OGD volume was determined from the total number of OGDs performed during the study period at each individual provider and separated into tertiles for analysis. The percentage of patients with UGIC diagnosed as an emergency was the ratio of UGIC subjects diagnosed by each provider during an acute (unplanned) admission divided by all UGIC subjects, including subjects diagnosed during an elective episode. The percentage

of patients who underwent a follow up OGD after a diagnosis of gastric ulcer was calculated by the number of patients with gastric ulcer coded at OGD who had a further OGD examination within 12 weeks divided by the total number of OGD coded with gastric ulcer. The provider PCCRC rates were determined by the proportion of CRC subjects who had a negative colonoscopy prior to their eventual CRC diagnosis. The gastric ulcer follow up percentage, percentage of UGIC as an emergency and provider PCCRC rates were analysed with providers separated into high, medium and low performing providers.

3.2.13 Provider variables for PCCRC study (Chapter 7)

For the purpose of this chapter, all endoscopy units operating within the same NHS organisation were analysed as a single provider. Individual providers were stratified by colonoscopy volume, BCSP status and the percentage of CRC diagnosed during an emergency rather than an elective episode to determine if there was an association with PCCRC. Colonoscopy volume was determined from the total number of colonoscopies performed during the study period at each provider and separated into tertiles. A BCSP accredited provider had at least one endoscopy unit accredited with BCSP status by the end of the study period. The percentage of CRC diagnosed as an emergency at a provider was the ratio of CRC diagnosed during an acute (unplanned) admission divided by all CRC, including CRC diagnosed during an elective episode.

3.2.14 Statistical methodology

Statistical analysis was carried out with STATA IC v12.1 (Statacorp LP, Texas, USA). Analysis of variance and $\chi 2$ tests were used to compare differences in continuous and categorical

variables respectively. Associations with POUGIC and PCCRC were examined by univariate and multivariate logistic regression analysis. For tests of significance p values <0.05 were considered significant. All odds ratios, 95% confidence intervals and associated p values are the result of multivariate analysis unless stated otherwise. Unadjusted Kaplan-Meier analysis and Cox proportional hazards model following adjustment for gender, age, deprivation and comorbidity was used to compare survival between the different groups studied.

A multivariate logistic regression analysis model was constructed to determine the association of subjects and OGD characteristics with POUGIC following adjusting for confounding factors including gender, age, Charlson comorbidity index, procedure type (emergency or elective), UGIC site (OC or GC), OC site (proximal, distal or unspecified), GC site (proximal, distal or unspecified), metastases and procedure year. Separate multivariate models were devised to adjust for the above variables for each of the provider variables (OGD volume, the percentage of UGIC patients diagnosed by an emergency procedure, the percentage of patients underwent following up OGD following diagnosis of gastric ulcer within 12 weeks and the provider PCCRC rates) to determine if there was an association with POUGIC.

A multivariate model was also constructed to determine associations with PCCRC following adjusting gender, age, Charlson co-morbidity index, procedure type (emergency or elective), CRC site (left colon or right colon), metastases and procedure year.

3.2.15 HES methodology and Structured Query Language (SQL) codes

The SQL codes used to extract data from HES for chapter 5 and chapter 7 are listed in appendix A and appendix B respectively.

3.2.16 Ethics

As only pseduonymised information was used in this study, ethics approval was not necessary. HES data is available under a data sharing agreement for the purposes of service evaluation.

3.3 West Midlands Cancer Intelligence Unit

(www.gov.uk/government/organisations/public-health-england)

Cancer registries are responsible for cancer registration that has been an integral part of the NHS for over 50 years. The national NHS contract mandates NHS providers to collect the cancer registration dataset and provide it to their local cancer registry. The West Midlands Cancer Intelligence Unit (WMCIU) which recently evolved into the National Cancer Registration Service (NCRS) West Midlands and is part of Public Health England in 2015.[48] The WMCIU covered a large geographic area including the counties of West Midlands, Worcestershire, Staffordshire, Herefordshire, Warwickshire and Shropshire. All 19 acute NHS trusts within the region are required to submit details of cancer cases diagnosed either as inpatients or outpatients under their care to WMCIU. The WMCIU also receives information from other sources such as general practitioners, community hospitals, hospices, pathology laboratories, and private hospitals. WMCIU also receives information

from death certificates if any diagnosis of cancer is recorded. WMCIU process this data and provide information to the Office of National Statistics. This national register is vital to understanding the trends in cancer incidence, prevalence and survival over time. This register also helps to understand trends within different geographical areas of the UK and in different social groups within the population. The data collected is used to support public health, healthcare policies and research.

3.3.1 Study design

A retrospective case-control study was carried out examining the POUGIC frequency and associated factor in the Midlands. NHS hospital trusts in the region were invited to take part with 11 trusts agreed to participate in the study. Following Information governance approval from each individual trust under a data sharing agreement (appendix 1) for the purposes of service evaluation, electronic endoscopy records were extracted from Burton Hospitals NHS Foundation Trust, The Dudley Group NHS Foundation Trust, George Eliot Hospital NHS Trust, Wye Valley NHS Trust, The Royal Wolverhampton Hospitals NHS Trust, Sandwell and West Birmingham Hospitals NHS Trust, South Warwickshire NHS Foundation Trust, University Hospital Birmingham NHS Foundation Trust, University Hospitals Coventry & Warwickshire NHS Trust, Worcestershire Acute Hospitals NHS Trust and Derby Teaching Hospitals NHS Foundation Trust. A limited dataset (surname, date of birth, NHS number and date of OGDs) of all OGD records extracted were submitted to WMCIU for cancer registration linkage. The overall POUGIC rate was obtained by dividing the number of POUGIC subjects by the number of all UGIC subjects.

3.3.2 WMCIU cancer registration linkage

Subjects who had an OGD record submitted to WMCIU were cross referenced with WMCIU registry database using at least two identification (name, NHS number and/or date of birth). Subjects with OC (ICD-10 code C15) or GC (ICD-1o code C16) registration were matched. Additional data for matched UGIC subjects were populated by WMCIU with cancer origin, tumour morphology (using the International Classification of Diseases for Oncology third edition (ICD-O-3) codes), UGIC diagnosis date, treatment methods and date of death if applicable.

3.3.3 Subject definition

POUGIC subjects were defined as any subjects having undergone OGD between 90 days to 3 years which did not detect UGIC prior to their eventual OC or GC diagnosis. Controls were subjects without having undergone OGDs within 3 years prior to UGIC diagnosis, except within 90 days prior to their UGIC diagnosis. Allowing a period of 90 days will ensure subjects with appropriate OGD follow up (such as checking for gastric ulcer healing) were not classified as POUGIC cases. Subjects with in situ carcinoma of the oesophagus or stomach, subjects who were on Barrett's surveillance program and subjects who had not undergone OGD within 3 years of UGIC diagnosis were excluded from the analysis.

3.3.4 Subject variables

Subject demographics including age and gender were compared between the POUGIC subjects and control group to investigation the association with POUGIC. UGIC variables including OC and GC, OC location (upper and mid oesophagus; lower oesophagus and

unknown or overlapping sites), GC location (fundus and cardia; gastric body; antrum and pylorus and unknown or overlapping sites), OC histology (adenocarcinoma; oesophageal squamous cell carcinoma; other histology types and unknown histology types), GC histology (diffused type adenocarcinoma; intestinal type adenocarcinoma; adenocarcinoma type not specified and unknown or other histology types) were also investigated for association with POUGIC. Indications for OGD were also extracted from endoscopy records with presenting symptoms of haematemesis, melanea, abdominal pain, anaemia, GORD, odynophagia, weight loss, nausea and vomiting, dysphagia, abnormal radiology and ulcer healing (for GC subjects only) were compared with subjects without these symptoms. Subjects with any red flag symptoms (haematemesis, melanea, anaemia, weight loss, nausea and vomiting and dysphagia) were also investigated for association with POUGIC.

3.3.5 Endoscopy variables

Pre-endoscopy medication of local anaesthesia (lidocaine throat spray), intravenous sedation (midazolam) and subjects received both local anaesthesia and intravenous sedation were investigated for association of POIGIC. The dosage of midazolam in POUGIC subjects and controls who received intravenous sedation was also compared with subjects separated into aged under 70 and over 70 groups. Endoscopist specialties were classified into medical endoscopists, surgical endoscopists, others (including nurse endoscopists) and endoscopists of unknown specialty. Trainee involvement (in which a second endoscopist was listed on the OGD report) was also interrogated to determine if this increase risk of associating with POUGIC. Endoscopic findings reported at OGD which did not detect UGIC (termed 'POUGIC OPD') were extracted for both OC and GC POUGIC subjects. The number

of biopsy taken from areas where abnormalities were report at POUGIC OGD and later UGIC detected at the same area were compared between POUGIC and control groups.

3.3.6 Treatment and survival outcomes

Treatments including surgery, chemotherapy and radiotherapy in addition to survival at 1 year post UGIC diagnosis were compared between POUGIC subjects and controls using records from WMCIU.

3.3.6 Statistical methodology

Statistical analysis was carried out with SPSS v20.0 (IBM, New York, USA). Independent t-test and $\chi 2$ test were used to compare differences in continuous and categorical variables respectively. Unconditional logistic regression analysis was used to calculate odds ratios and 95% confidence intervals (CI) of the influence of age, gender, UGIC type (OC or GC), UGIC location, UGIC histology type, presenting upper gastrointestinal symptoms, presence of red flag symptoms, pre-procedure medication use, intravenous midazolam dosage, endoscopist specialty, trainee involvement, biopsy numbers from POUGIC ODG, surgery, chemotherapy, radiotherapy and survival at 1 year on POUGIC. For tests of significance p values <0.05 were considered significant.

3.3.7 Ethics

The project protocol was reviewed by the National Research Ethics Service (NRES) who advised as the project intent fall in the clinical audit category, NRES review and ethics

approval was not required. This was based on the primary aim being to measure standards of care and assessing the current clinical practice.

Chapter 4: How common is upper gastrointestinal cancer occurrence after endoscopy that did not diagnose cancer (A UK primary care based study)?

(Data from this chapter has been published: **Cheung D, Menon S, Hoare J, Dhar A, Trudgill N. Factors Associated with Upper Gastrointestinal Cancer Occurrence After Endoscopy that Did Not Diagnose Cancer.** ***Digestive diseases and sciences*** **2016;61:2674-84**.)

4.1 Introduction

The prognosis of UGIC in the UK is extremely poor, with 5 year survival rates for OC and GC of 12% and 16% respectively.[44] The poor prognosis of UGIC relates to it usually presenting at an advanced stage, with only one third of UGIC subjects suitable for curative treatment.[49] The prognosis for subjects with early stage disease, who are eligible for curative resection, has improved [50, 51] with 80% alive at 1 year.[52] However, the prognosis for subjects with locally advanced and metastatic disease remains poor. Selected single institution studies in Western populations in a total of 908 subjects have reported that 4.6% to 14.0% of UGIC subjects have had an OGD which did not identify UGIC in the 3 years prior to their eventual UGIC diagnosis.[12, 13, 14, 15, 16] These events termed POUGIC following the same principle as PCCRC.[53] Subjects who presented with alarm symptoms, including dysphagia, anaemia, hematemesis, weight loss or vomiting, at the time of OGD have been reported to be at increased risk of POUGIC.[13, 14] In addition, squamous cell carcinoma in the proximal oesophagus[14] and taking less biopsy specimens[13, 15] were reported to be associated with POUGIC. Subject characteristics such as age and gender did not appear to affect the likelihood of POUGIC.[13, 14, 15, 16]

4.2 Materials and methods

Using THIN data, a UK primary care dataset, the POUGIC rate was determined at a national level in an unselected sample and associated risk factors for these events were identified. THIN data between 2002 and 2009 (for GC subjects) and 2002 and 2012 (for OC subjects) were used in this study. The treatment and survival outcomes for POUGIC subjects were also studied.

4.3 Results

There were 11966 UGIC subjects during the study period, with 5473 GC and 6493 OC subjects. Following exclusion of subjects who did not meet the study criteria, 4249 GC (44.8%) and 5238 OC subjects (55.2%) were included for analysis.

4.3.1 Subject characteristics

The POUGIC subject characteristics are shown in Table 4.1. There were 633 POUGIC subjects, 279 with OC and 354 with GC. The overall POUGIC rate was 6.7%, with the POUGIC rate for OC and GC being 5.3% and 8.3% respectively. POUGIC subjects were more likely to have GC than OC. This was less marked when adjusted for other variables but remained a significant association. Younger age and female gender were associated with POUGIC. When UGIC subjects were separated into OC and GC subjects, the age association was only observed in GC subjects and the female gender association was only observed in OC subjects. Increasing medical co-morbidity was associated with POUGIC. Subjects with a Charlson co-morbidity score of 1 to 4 were at modestly increased risk compared with subjects without co-morbid illnesses in univariate and multivariate analyses. Increasing

deprivation was associated with POUGIC, with more deprived postcodes (Townsend score 4th and 5th quintiles) more likely to be associated with POUGIC compared with Townsend score 1st and 2nd quintiles. This association remained statistically significant following adjusting for confounding factors.

4.3.2 Presenting symptoms prior to UGIC diagnosis

Presenting symptoms prior to UGIC diagnosis are shown in Table 4.2. Subjects who presented with alarm symptoms within 12 months of their OGD were much less likely to be associated with POUGIC. This effect was even more notable in subjects with OC compared with subjects with GC. Alarm symptoms remained strongly associated even after adjusting for potential confounding factors.

In subjects with OC, POUGIC subjects were most likely to present with GORD symptoms (45.2%), whereas controls were mostly likely to present with dysphagia (44.8%) in the 12 months prior to their POUGIC OGD and diagnostic OGD respectively. OC subjects who presented with dysphagia, weight loss or vomiting were all less likely to be associated with POUGIC. In contrast, OC subjects with GORD symptoms were nearly three times more likely to be associated with POUGIC.

In subjects with GC, both POUGIC subjects (40.1%) and controls (20.4%) were more likely to present with GORD symptoms. However, presenting with GORD symptoms increased the risk of GC POUGIC more than two fold. Symptoms of anaemia, vomiting, weight loss, dysphagia or anorexia were all negatively associated with POUGIC in GC subjects.

Table 4.1. The subject characteristics of post-OGD upper gastrointestinal cancer cases and upper gastro-intestinal cancer controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Number of subjects				**Univariate**			**Multivariate**		
Total	633 (6.7%)	8854 (93.3%)	9487						
OC	279 (5.3%)	4959 (94.7%)	5238	Ref	-	-	Ref	-	-
GC	354 (8.3%)	3895 (91.7%)	4249	**1.62**	**1.37-1.90**	**<0.0001**	**1.33**	**1.13-1.58**	**0.001**
Mean age (years±SD)									
Total	70.2±11.2	72.8±11.2			**1.62-3.43**	**<0.0001**	**1.02**	**1.01-1.03**	**<0.0001**
OC	70.5±10.8	71.8±11.4			0.07-2.66	0.064			
GC	70.1±11.6	74.1±11.0			2.79-5.18	<0.0001			
Gender									
UGIC male	371 (6.0%)	5766 (94.0%)	6137	Ref	-	-	Ref	-	-
UGIC female	262 (7.8%)	3088 (92.2%)	3350	**1.31**	**1.12-1.55**	**0.001**	**1.39**	**1.17-1.64**	**<0.0001**
OC male	159 (4.6%)	3310 (95.4%)	3469	Ref	Ref	Ref			
OC female	120 (6.8%)	1649 (93.2%)	1769	**1.51**	**1.19-1.93**	**0.0009**			
GC male	212 (7.9%)	2456 (92.1%)	2668	Ref	Ref	Ref			
GC female	142 (9.0%)	1439 (91.0%)	1581	1.14	0.92-1.43	0.238			
Charlson comorbidity index									
0	224 (5.8%)	3659 (94.2%)	3883	Ref	-	-	Ref	-	-
1 to 4	381 (7.2%)	4899 (92.8%)	5280	**1.27**	**1.07-1.51**	**0.006**	**1.35**	**1.13-1.61**	**0.001**
5 or greater	28 (8.6%)	296 (91.4%)	324	**1.55**	**1.03-2.33**	**0.038**	1.41	0.92-2.14	0.113
Deprivation by Townsend score quintile									
1st and 2nd	226 (5.7%)	3761 (94.3%)	3987	Ref	-	-	Ref	-	-
3rd	130 (6.9%)	1748 (93.1%)	1878	1.24	0.99-1.55	0.061	1.24	0.99-1.56	0.06
4th and 5th	239 (7.6%)	2887 (92.4%)	3126	**1.38**	**1.14-1.66**	**0.0008**	**1.31**	**1.09-1.59**	**0.005**
Unknown	38 (7.7%)	458 (92.3%)	496	1.38	0.97-1.97	0.077	1.23	0.86-1.77	0.260

Table 4.2. Consultations with upper gastrointestinal symptoms in the 12 months prior to post-OGD upper gastrointestinal cancer endoscopy and prior to upper gastrointestinal cancer diagnosis in controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Alarm symptoms within 12 months of POUGIC OGD and diagnostic OGD for controls				**Univariate**			**Multivariate**		
Total	126 (19.9%)	3982 (45.0%)	4108	**0.30**	**0.25-0.37**	**<0.0001**	**0.32**	**0.26-0.40**	**<0.0001**
OC	62 (22.2%)	2650 (53.4%)	2712	**0.25**	**0.19-0.33**	**<0.0001**			
GC	64 (18.1%)	1332 (34.2%)	1396	**0.42**	**0.32-0.56**	**<0.0001**			
Symptoms within 12 months of POUGIC OGD and diagnostic OGD for controls (OC subjects)									
Abdominal mass	2 (0.7%)	22 (0.4%)	24	1.62	0.38-6.93	0.515			
Anaemia	18 (6.5%)	225 (4.5%)	243	1.45	0.88-2.38	0.141			
Anorexia	4 (1.4%)	102 (2.1%)	106	0.69	0.25-1.89	0.474			
Dysphagia	27 (9.7%)	2220 (44.8%)	2247	**0.11**	**0.07-0.16**	**<0.0001**			
Haematemesis/ melaena	11 (3.9%)	120 (2.4%)	131	1.66	0.88-3.11	0.117			
GORD symptoms	126 (45.2%)	1136 (22.9%)	1262	**2.77**	**2.17-3.54**	**<0.0001**			
Vomiting	10 (3.6%)	356 (7.2%)	366	**0.48**	**0.25-0.91**	**0.025**			
Weight loss	5 (1.8%)	214 (4.3%)	219	**0.40**	**0.17-0.99**	**0.048**			
Symptoms within 12 months of POUGIC OGD and diagnostic OGD for controls (GC subjects)									
Abdominal mass	5 (1.4%)	64 (1.6%)	69	0.86	0.34-2.14	0.743			
Anaemia	14 (4.0%)	477 (12.2%)	491	**0.30**	**0.17-0.51**	**<0.0001**			
Anorexia	7 (2.0%)	131 (3.4%)	138	0.58	0.27-1.25	0.164			
Dysphagia	18 (5.1%)	371 (9.5%)	389	**0.51**	**0.31-0.83**	**0.006**			
Haematemesis/ melaena	19 (5.4%)	288 (7.4%)	307	0.71	0.44-1.15	0.160			
GORD symptoms	142 (40.1%)	793 (20.4%)	935	**2.62**	**2.09-3.29**	**<0.0001**			
Vomiting	14 (4.0%)	371 (9.5%)	385	**0.39**	**0.23-0.67**	**0.0007**			
Weight loss	7 (2.0%)	182 (4.7%)	189	**0.41**	**0.19-0.88**	**0.023**			

4.334 Endoscopic findings

The endoscopic findings from POUGIC OGDs are shown in Table 4.3. The most common finding was oesophagitis in 19.4% of POUGIC subjects with OC and gastritis in 22.6% of POUGIC subjects with GC. Endoscopic findings recognised to be associated with OC (oesophageal stricture and ulcer) were reported in 5.7% of OC POUGIC cases and findings associated with GC (gastric ulcer) were reported in 10.5% of GC POUGIC cases. Of the POUGIC subjects with OC who had an oesophageal stricture or ulcer reported at POUGIC OGD and POUGIC subjects with GC who had a gastric ulcer reported at POUGIC OGD, only 50.0% and 64.6% respectively had a follow up OGD within 90 days. POUGIC subjects who presented with alarm symptoms were significantly more likely to have oesophageal stricture and gastric ulcer reported at their POUGIC OGD.

Subjects with OC in the lower oesophagus appeared to be at lower risk of POUGIC compared with subjects with OC in the upper and mid oesophagus but there was no significant association, in part due to the large number of subjects with unknown UGIC location (Table 4.4). There was no difference in the site of GC in POUGIC subjects, with equal proportions of proximal and distal GC in POUGIC subjects and controls.

Table 4.3. Endoscopic findings at post-OGD upper gastrointestinal cancer endoscopy

OGD findings in POUGIC subjects with OC	With alarm symptoms (n=49)	Without alarm symptoms (n=230)	Total	Odds ratio	95% CI	p value
Oesophageal stricture	7	5	12 (4.3%)	**7.50**	**2.27-24.75**	**0.0009**
Oesophageal ulcer	0	4	4 (1.4%)	0.51	0.03-9.60	0.652
Oesophagitis	6	48	54 (19.4%)	0.53	0.21-1.32	0.171
Gastritis	5	35	40 (14.3%)	0.63	0.23-1.71	0.367
Gastric ulcer	1	6	7 (2.5%)	0.78	0.09-6.61	0.818
Duodenitis	1	8	9 (3.2%)	0.58	0.07-4.73	0.609
Duodenal ulcer	4	10	14 (5.0%)	1.96	0.59-6.51	0.275
OGD findings in POUGIC subjects with GC	With alarm symptoms (n=49)	Without alarm symptoms (n=305)				
Oesophageal stricture	2	1	3 (0.8%)	**12.94**	**1.15-145.49**	**0.038**
Oesophageal ulcer	0	1	1 (0.3%)	2.05	0.08-51.05	0.662
Oesophagitis	4	42	46 (13.0%)	0.56	0.19-1.63	0.285
Gastritis	10	70	80 (22.6%)	0.86	0.41-1.81	0.693
Gastric ulcer	10	27	37 (10.5%)	**2.64**	**1.19-5.87**	**0.017**
Duodenitis	1	7	8 (2.3%)	0.89	0.11-7.37	0.912
Duodenal ulcer	1	24	25 (7.1%)	0.24	0.03-1.85	0.172

Table 4.4. Site of oesophageal and gastric cancers

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
				Univariate		
Upper/mid OC	16 (5.7%)	184 (3.7%)	200	Ref	-	-
Lower OC	41 (14.7%)	799 (16.1%)	840	0.59	0.32-1.07	0.085
OC unknown	222 (79.6%)	3976 (80.2%)	4198	0.64	0.38-1.09	0.100
Proximal GC	31 (8.8%)	329 (8.4%)	360	Ref	-	-
GC body	3 (0.8%)	77 (2.0%)	80	0.41	0.12-1.39	0.153
Distal GC	18 (5.1%)	158 (4.1%)	176	1.21	0.66-2.23	0.543
GC unknown	302 (85.3%)	3331 (85.5%)	3633	0.96	0.65-1.42	0.845

4.3.4 UGIC treatment outcomes and survival

The UGIC treatment outcomes and survival are shown in Tables 4.5, 4.6 and 4.7. POUGIC subjects were more likely to undergo surgery than controls on univariate analysis. However, this association was confined to male subjects with GC. There was no difference between POUGIC subjects and controls undergoing chemotherapy. However, when separating subjects with OC and GC by gender, female POUGIC subjects, POUGIC subjects with OC and particularly female POUGIC subjects with OC were more likely to have chemotherapy. In contrast, male POUGIC subjects with GC were less likely to undergo chemotherapy. Following adjusting for confounding factors, POUGIC subjects were marginally more likely to undergo radiotherapy compared with controls but there was no overall difference in the likelihood of undergoing surgery or chemotherapy. When comparing POUGIC subjects with controls, there was no difference in 1 year survival and overall survival (figure 4.1). When sub-analysis was carried out by separating subjects with OC and GC, POUGIC subjects with GC were more likely to survive at 1 year compared with controls.

Table 4.5. Treatment outcomes and adjusted survival for post-OGD upper gastrointestinal cancer subjects and upper gastrointestinal cancer controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
UGIC subjects undergoing surgery				**Univariate**			**Multivariate**		
All	150 (23.7%)	1636 (18.5%)	1786	**1.37**	**1.13-1.66**	**0.001**	1.19	0.98-1.46	0.082
Male	99 (26.7%)	1151 (20.0%)	1250	**1.46**	**1.15-1.85**	**0.002**			
Female	51 (19.5%)	485 (15.7%)	536	1.30	0.94-1.79	0.112			
UGIC subjects undergoing chemotherapy									
All	146 (23.1%)	2035 (23.0%)	2181	1.00	0.83-1.22	0.963	1.193	0.98-1.46	0.087
Male	83 (22.4%)	1478 (25.6%)	1561	0.84	0.65-1.07	0.163			
Female	63 (24.0%)	557 (18.0%)	620	**1.44**	**1.07-1.94**	**0.017**			
UGIC subject undergoing radiotherapy									
All	58 (9.2%)	704 (8.0%)	762	1.17	0.88-1.55	0.279	**1.38**	**1.03-1.84**	**0.029**
Male	33 (8.9%)	455 (7.9%)	488	1.14	0.79-1.65	0.489			
Female	25 (9.5%)	249 (8.1%)	274	1.20	0.78-1.85	0.402			
Survival at 1 year for UGIC subjects									
All	219 (34.6%)	2820 (31.9%)	3039	1.13	0.95-1.34	0.153	1.06	0.89-1.27	0.496
Male	130 (35.0%)	1895 (32.9%)	2025	1.10	0.88-1.37	0.388			
Female	89 (34.0%)	925 (30.0%)	1014	1.20	0.92-1.57	0.175			

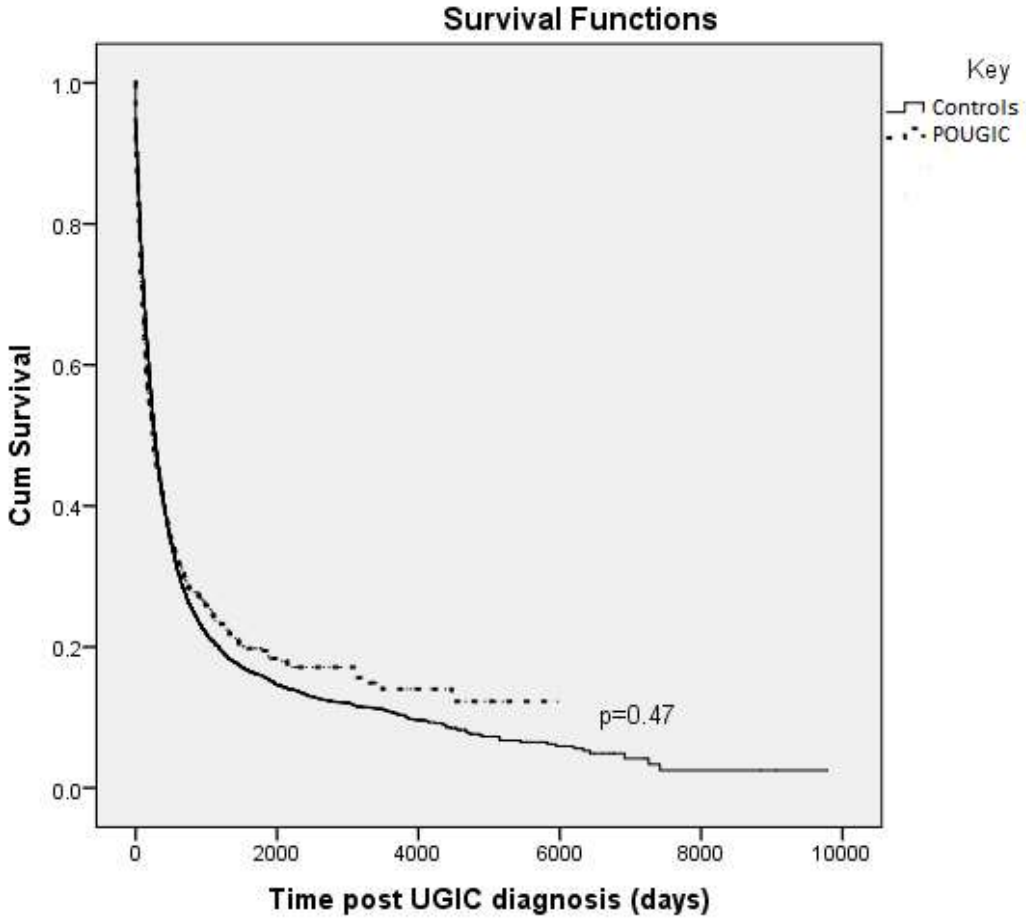
Table 4.6. Treatment outcomes and unadjusted survival for post-OGD upper gastrointestinal cancer subjects with oesophageal cancer and oesophageal cancer controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
OC subjects undergoing surgery				**Univariate**		
All	53 (19.0%)	818 (16.5%)	871	1.19	0.87-1.62	0.276
Male	38 (23.9%)	624 (18.9%)	662	1.35	0.93-1.97	0.115
Female	15 (12.5%)	194 (11.8%)	209	1.07	0.61-1.88	0.810
OC subjects undergoing chemotherapy						
All	81 (29.0%)	1172 (23.6%)	1253	**1.32**	**1.01-1.73**	**0.040**
Male	42 (26.4%)	825 (24.9%)	867	1.08	0.75-1.55	0.672
Female	39 (32.5%)	347 (21.0%)	386	**1.81**	**1.21-2.70**	**0.004**
OC subject undergoing radiotherapy						
All	34 (12.2%)	554 (11.2%)	588	1.10	0.76-1.60	0.602
Male	18 (11.3%)	357 (10.8%)	375	1.06	0.64-1.75	0.832
Female	16 (13.3%)	197 (11.9%)	213	1.13	0.66-1.96	0.652
Survival at 1 year for OC subjects						
All	102 (36.6%)	1734 (35.0%)	1836	1.07	0.83-1.38	0.588
Male	60 (37.7%)	1185 (35.8%)	1245	1.09	0.78-1.51	0.619
Female	42 (35.0%)	549 (33.3%)	591	1.08	0.73-1.59	0.702

Table 4.7. Treatment outcomes and unadjusted survival for post-OGD upper gastrointestinal cancer subjects with gastric cancer and gastric cancer controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
GC subjects undergoing surgery						
All	97 (27.4%)	818 (21.0%)	915	**1.42**	**1.11-1.82**	**0.005**
Male	61 (28.8%)	527 (21.5%)	588	**1.48**	**1.08-2.02**	**0.014**
Female	36 (25.4%)	291 (20.2%)	327	1.34	0.90-2.00	0.151
GC subjects undergoing chemotherapy						
All	65 (18.4%)	863 (22.2%)	928	0.79	0.60-1.05	0.987
Male	41 (19.3%)	653 (26.6%)	694	**0.66**	**0.47-0.94**	**0.022**
Female	24 (16.9%)	210 (14.6%)	234	1.19	0.75-1.89	0.461
GC subject undergoing radiotherapy						
All	24 (6.8%)	150 (3.9%)	174	**1.82**	**1.16-2.83**	**0.009**
Male	15 (7.1%)	98 (4.0%)	113	**1.83**	**1.04-3.22**	**0.035**
Female	9 (6.3%)	52 (3.6%)	61	1.80	0.87-3.74	0.113
Survival at 1 year for GC subjects						
All	117 (33.1%)	1086 (27.9%)	1203	**1.28**	**1.01-1.61**	**0.039**
Male	70 (33.0%)	710 (28.9%)	780	1.21	0.90-1.64	0.207
Female	47 (33.1%)	376 (26.1%)	423	1.40	0.97-2.02	0.075

Figure 4.1 Unadjusted overall survival between POUGIC subjects and controls



4.3.5 Change in POUGIC incidence with time

OC subjects undergoing OGD prior to 2008 were between 2 to 3 times more likely to be associated with POUGIC than subjects undergoing OGD after 2008 ($p<0.0001$, $p=0.0001$)(Table 4.8). The difference with time period was less marked in subjects with GC, with subjects undergoing OGD prior to 2005 1.5 times more likely to have POUGIC, compared with subjects undergoing OGD after 2005 ($p=0.014$, $p=0.003$).

Table 4.8. The frequency of post-OGD Upper gastrointestinal cancer by time period

POUGIC or diagnostic OGD date for controls	POUGIC	Controls	Total	Odds ratio	95% CI	p value
OC				**Univariate**		
Prior to 2004	138 (7.9%)	1608	1746	**3.10**	**2.21-4.35**	**<0.0001**
2004 to 2008	94 (5.4%)	1652	1746	**2.06**	**1.44-2.94**	**0.0001**
2008 to 2012	47 (2.7%)	1699	1746	Ref	-	-
GC						
Prior to 1999	127 (9.0%)	1289	1416	**1.42**	**1.07-1.88**	**0.014**
1999 to 2005	135 (9.5%)	1281	1416	**1.52**	**1.15-2.00**	**0.003**
2005 to 2011	92 (6.5%)	1325	1417	Ref	-	-

4.4 Discussion

OGD is the gold standard for investigating upper gastrointestinal symptoms and diagnosing UGIC. In a recent meta-analysis, POUGIC was found to be relatively uncommon occurring in approximately 1 in every 400 OGDs.[26] However, POUGIC was relatively common among UGIC subjects, with 4.6% to 14.0% having had an OGD which did not detect UGIC in the preceding 3 years [12, 14, 15, 16, 23]. Overall, POUGIC occurs in 6.4% of UGIC subjects within 1 year of diagnosis and in 11.3% of UGIC subjects up to 3 years before diagnosis.[26] Two recent population based UK studies have reported that 8.3% of GC and 7.7% of OC subjects have had an OGD up to 3 years prior to eventual UGIC diagnosis.[24, 25] An interval of 3 years is derived from the assumption that the doubling time for mucosal GC is 2 to 3 years from a Japanese study from the 1970's[20], and this interval is commonly used to define a false-negative endoscopic examination in the detection of UGIC. The POUGIC rate from this study, the largest ever of this issue, was 6.7%.

In the current study, younger age and female gender were more likely to be associated with POUGIC. Similar findings have been reported in a recent UK series based on a national gastric cancer audit.[25] This could potentially be explained by younger subjects [54, 55, 56] and women [55, 57] reportedly having a lower tolerance for OGD examination, which may in turn lead to a reduction in OGD diagnostic quality. Another possible explanation might be the lower expectation of UGIC in women and younger subjects by endoscopists, due to the lower incidence of UGIC in younger and female subjects. The increased risk of POUGIC in women in the present study was only related to OC and not GC. Squamous cell OC accounts for 65.4% of all OC in women but only 28.6% in men, in whom oesophageal adenocarcinoma is much more common.[58] Unlike the readily recognisable signs of early oesophageal adenocarcinoma such as Barrett's oesophagus, the early signs of squamous cell OC may be less readily recognised in Western populations.[14, 59] This may explain the gender difference found in the current study.

Subjects with increasing medical co-morbidity were more likely to have an episode of POUGIC, which might relate to a lower tolerance of the procedure due to their associated medical conditions and therefore quality of OGD examination. Alternatively, subjects with multiple co-morbidities may be more likely to undergo OGD than subjects without co-morbidity for conditions such as anaemia related to their co-morbidities, when a relatively small, asymptomatic early UGIC might not be detected.

UGIC subjects who presented with alarm symptoms within 12 months of their OGD were much less likely to be associated with POUGIC. Alarm symptoms suggest a more advanced case of UGIC, and thus the UGIC would be more likely to be detected during OGD

examination. In contrast, presenting with haematemesis or melaena or GORD symptoms is not usually associated with UGIC and therefore this may potentially affect the endoscopist's awareness of early UGIC during OGD. Surprisingly, the opposite finding to this has been reported in series in Scotland and Western Australia with subjects presenting with alarm symptoms being more likely to experience POUGIC.[13, 14] The difference with the findings from these studies is likely to relate to identifying POUGIC subjects within six months of UGIC diagnosis, rather than 12 months in the present study, and diagnosis being delayed in advanced UGIC cases due to food residue or blood obscuring the view, inadequate biopsy sampling or follow up arrangements. In the present study, POUGIC subjects who presented with alarm symptoms were more likely to have endoscopic findings such as oesophageal stricture and gastric ulcer reported at their POUGIC OGD that are known to be associated with UGIC. Such endoscopic lesions were reported in up to 8.3% of POUGIC cases. Of these, only 50.0% of subjects with oesophageal stricture or ulcer and 64.6% of subjects with gastric ulcer had a follow up OGD within 90 days in the current study. A lack of adequate follow up of these lesions is likely to be a contributing factor to POUGIC cases. The current JAG standard stipulates repeat OGD for all gastric ulcers within 12 weeks. It was unclear if the cases which were not followed up were due to administrative error, service capacity limitations or clinical decision.

POUGIC subjects appeared more likely to undergo surgery following UGIC diagnosis, however, this was likely to be due to confounding factors (such as younger age) as there was no association after adjusting for other variables. Overall, there was no difference in both unadjusted and adjusted survival at 1 year between POUGIC subjects and controls. The

same findings were also reported in a Finnish and a recent UK cohort.[16, 24, 25] This should not be surprising given the very poor overall survival in UGIC patients and obviously the situation might potentially be very different if the POUGIC had been diagnosed at an earlier opportunity.

Encouragingly, the POUGIC rate in the UK has fallen over the study period from 7.9% to 2.7% for OC and 9.0% to 6.5% for GC. There are likely to be a number of factors behind this fall including improvements in endoscopic pathways, such as routinely following up oesophageal strictures or ulcers and gastric ulcers (which has improved from 55.9% to 69.8% when comparing periods before and after 2000 in the dataset), endoscopists taking more biopsies from suspicious lesions, improvements in the quality of endoscopic imaging and endoscopists becoming more aware of early signs of UGIC. The reasons for POUGIC being more commonly associated with GC than OC cannot be identified in the present study. The oesophagus has a smaller surface area, simpler anatomy and the mucosa is less likely to be contaminated than the stomach with food, debris or bile impeding the endoscopic view. Endoscopists in the UK are also likely to be less aware of gastric pre-malignant changes such as gastric atrophy or intestinal metaplasia than the more widely recognised pre-malignant condition Barrett's oesophagus and this may contribute to more early GC than early OC not being recognised at OGD.

The large sample size and its unselected nature are the obvious strengths of the present study, making it the largest study on POUGIC to date. The total of 9487 UGIC subjects included was greater than the sum of all subjects included in previous studies of POUGIC. The THIN database spans over two decades, allowing changes in POUGIC incidence to be

examined. The THIN primary care centres are spread across the UK and subjects are regionally and demographically representative of the UK. In addition, as patients must be registered with a primary care practitioner in order to access secondary care services, this allowed unbiased subject selection, which is a potential source of bias in most previous studies due to their subject cohorts being recruited from a single healthcare provider. Furthermore, the data captured in THIN has previously been validated in a number of studies.[60, 61]

Despite the above advantages, there are a number of limitations including specific issues related to the THIN dataset. The lack of ability to link the THIN dataset to the National cancer registry data is a significant disadvantage. However, primary care practitioners contributing data to THIN follow a standardised process and codes for cancer would not be entered without histological confirmation from a secondary care provider. In order to further validate the dataset in the current study, the surgical rates for OC (14.5%) and GC (15.1%) in 2010 from THIN were compared with the National Oesophagogastic cancer audit. The national audit reported a surgical rate of 20.0% in OC subjects and 22.4% in GC subjects respectively during the same period with a case-ascertainment of 71.1%.[46] Furthermore, the 1 year survival rate in the present study was similar to national survival rates reported in cancer registry data. In THIN, the survival rate for OC subjects diagnosed between 1997 to 1999 was 36.1% and subjects diagnosed between 2000 to 2002 was 33.8%, which is comparable to cancer registry rates of 33.3% and 38.0% respectively.[62] The possibility of administrative delays in primary care in recording the UGIC diagnosis date, led us to exclude the period within 12 months of UGIC diagnoses for analysis of POUGIC, potentially excluding some POUGIC cases. However, although addressing the reasons for patients undergoing an

OGD that did not diagnose UGIC within a few months of their diagnosis is an important issue, it is much less likely to improve the prognosis of UGIC than diagnosing the UGIC at an earlier stage or as a pre-malignant lesion years before the diagnosis date in POUGIC cases. THIN only captures diagnostic outcomes from OGDs and data potentially relevant to POUGIC, such as whether sedation was used, the grade and specialty of the endoscopist, *Helicobacter pylori* status, if biopsies were taken and the number of biopsies taken are not recorded, limiting conclusions on why POUGIC cases occurred. Furthermore, the lack of complete data on UGIC histology and UGIC staging further limited analysis of potential causes of POUGIC and the degree to which an endoscopist could potentially be responsible for a case of POUGIC. For example, there may be virtually no changes at OGD three years before later diagnosis with an early stage UGIC, whereas thirteen months before presenting with an advanced UGIC it is very likely that the endoscopist missed an existing malignant lesion.

In summary, in the largest study to date, the risk of POUGIC among UGIC subjects was 6.7%. POUGIC was associated with younger age, female gender, increasing co-morbidity, increasing deprivation and a lack of alarm symptoms at presentation. POUGIC was more common among GC subjects. Endoscopic findings such as stricture and ulceration that are known to be associated with UGIC were recorded in 8.3% of POUGIC OGDs, representing potential missed opportunities for early UGIC diagnosis.

Chapter 5: Factors associated with upper gastro-intestinal cancer occurrence after endoscopy that did not diagnose upper gastro-intestinal cancer (an analysis of English Hospital Episode Statistics from 2003 to 2012).

5.1 Introduction

OGD is the investigation of choice for excluding UGIC consisting mainly of OC and GC. UGIC is the fifth most common malignancy in the UK, with over 15,000 new diagnosis and accounting for around 12,000 deaths per annum.[63] The number of patients diagnosed with UGIC have been relatively unchanged over the last decade, with the falling number of GC cases balanced by increasing OC incidence.[44] Currently, the UK has the highest age-standardised incidence of OC in Europe[64] with adenocarcinoma accounting for two thirds of all OC.[49]

The prognosis for patients diagnosed with early stage disease who are suitable for endoscopic resection or curative surgery has dramatically improved over the last few decades.[50, 51] However, UGIC commonly presents late with the majority receiving palliative and supportive management only.[65] The prognosis for patients with locally advanced disease and metastatic disease remains poor, with the overall 5 year survival rate of OG and GC at 12% and 16% respectively with no improvement in survival rate over the last decade.[4, 44] As prognosis for patients who are eligible for curative treatment are far superior, therefore, recognition and early diagnosis of UGIC is paramount.

There are very few auditable outcomes and quality standards for OGD in the UK at present with successful intubation, completeness of procedure (in reaching the distal duodenum) and follow up OGD for gastric ulcers healing within 12 weeks as the only recommendations set out by the JAG.[66] In comparison, colonoscopy has extensive standards set out a decade ago which has since experienced major improvements in procedure quality.[30] Furthermore, published studies suggest between 6.7% to 14.0% of UGIC subjects have had an OGD which did not identify OC or GC in the 3 years prior to UGIC diagnosis.[12, 13, 14, 15, 16] These POUGIC cases which encompass the two likely theories in which POUGIC are either UGIC was not detected on previous false negative OGD or interval cancer with aggressive tumour biology developed following previous negative OGD. Previous POUGIC studies on Western population have reported patients presented with alarm symptoms (including dysphagia, anaemia, haematemesis, weight loss or vomiting) at the time of OGD were at an increased risk of POUGIC.[13, 14] Squamous cell carcinoma in the proximal oesophagus[14] and taking few biopsy specimens[13, 15] were also reported to be associated with POUGIC.

5.2 Materials and methods

Most of the published studies to date were retrospective analysis of patient cohort from a single healthcare provider which the findings may be affected with unavoidable patient selection bias.

The frequency of POUGIC at national level and associated risk factors for these events were examined using a national hospital episode database in England. HES data between 2003 and 2012 were used for this study.

5.3 Results

Between 2003 and 2012, there were a total of 112,430 UGIC subjects identified from HES following excluding subjects who did not meet study criteria. There were 11,225 POUGIC subjects, of these, 4029 (35.9%) subjects have had a negative OGD between 6-12 months prior to eventual UGIC diagnosis and a further 7196 (64.1%) subjects who had a negative OGD between 12-36 months prior to UGIC diagnosis. The overall POUGIC during this period in England was 10.0%.

5.3.1 Validation of upper gastrointestinal cancer population studied

The total number of UGIC subjects identified from HES between 2008 and 2011 were 50,814 compared with total of 51,617 UGIC subjects reported in National Cancer Intelligence Network (NCIN). This equates to 98% of UGIC subjects were capture in HES for the same period. Additionally, the surgical and chemotherapy rates for subjects with UGIC from HES were compared with the National oesophagogastric cancer audit between 2002 and 2009. Where data was available the overall surgical rate and chemotherapy rate from NCIN was 21.1% and 45.6% respectively, this is broadly in keeping with the figures from HES (surgical rate 20.1% and chemotherapy rate of 35.0%). The discrepancy may be due to the cases ascertain rate of NCIN were around 70% and subjects undergoing treatment being more likely to be reported for auditing.

5.3.2 Subject characteristics

The POUGIC subject characteristics are shown in Table 5.1. POUGIC subjects were marginally younger than controls (by 0.3 years, p=0.003). When separated in age group tertiles, subjects in the oldest age tertile were less likely be associated with POUGIC when compared with subjects in the youngest age tertile after adjusting for confounding factors (0.88 (95% CI 0.81-0.95), p=0.001). POUGIC subjects were also more likely to be female (1.20 (95% CI 1.12-1.27), p<0.001). Subjects who has significant comorbidity (Charlson index 5 or greater) were protective of POUGIC (0.80 (95% CI 0.74-1.19), p<0.001). Subjects in least deprived quintiles were also less likely to be associated with POUGIC, when compared with subjects in the most deprived quintile. In terms of ethnicity, subjects with an Asian or Asian British background (1.37 (95% CI 1.18-1.59), p=0.001) and Black or Black British subjects (1.33 (95% CI 1.15-1.54), p=0.001) were both more likely to be associated with POUGIC then subjects with a White or White British background.

5.3.3 Endoscopic factors

Upper GI endoscopy characteristics and findings of POUGIC OGDs are shown in Table 5.2. Emergency OGDs which are usually carried out for inpatients were less likely to be associated with POUGIC (0.77 (95% CI 0.71-0.84), p<0.001) when compared with elective OGDs which typically performed as day case procedures. There appears to be no difference between OGDs carried out during the week or at the weekend (1.10 (95% CI 0.99-1.24), p=0.087). Subjects with GC were more likely to be associated with POUGIC than OC subjects (1.05 (95% CI 1.01-1.09), p=0.021). When analysing OC and GC subjects separately, OC subjects with distal oesophageal cancer were protective from POUGIC (0.87 (95% CI 0.80-

0.95), p=0.002) when compared with OC subjects with proximal oesophageal cancer. There were no difference when comparing mid stomach GC and distal stomach GC with proximal stomach GC.

Pathology detected at OGDs which did not detect UGIC in the POUGIC subjects ('POUGIC OGD') were examined with oesophageal ulcer present in 38.5% and oesophageal stricture in 6.6% in all POUGIC subjects with OC. Gastric ulcer was present in 26.1% of all POUGIC OGDs for subjects who eventually GC was detected at a later date.

5.3.4 Treatment and survival outcomes

The prevalence of metastases within 12 months of diagnosis and treatment outcomes are shown in Table 5.3. After adjusting for confounding factors, POUGIC subjects were less likely to have liver metastases (0.56 (95% CI 0.51-0.63), p<0.001), lung metastases (0.84 (95% CI 0.79-0.89), p<0.001), peritoneal metastases (0.75 (95% CI 0.59-0.95), p=0.017) and lymph node metastases (0.55 (95% CI 0.49-0.61), p<0.001) than controls within 12 months of UGIC diagnosis.

Following UGIC diagnosis, POUGIC subjects were less likely to undergo surgery (0.84 (95% CI 0.80-0.87), p<0.001) and half as likely to undergo chemotherapy (0.52 (95% CI 0.50-0.55), p<0.001) when compared with controls. There were no difference between POUGIC subjects and control with radiotherapy (0.97 (95% CI 0.90-1.04), p=0.418). POUGIC subjects had an unadjusted survival advantage compared with controls (figures 5.1, 5.2 and 5.3). This effect appears most marked between POUGIC subjects with OC and controls with OC.

Figure 5.1 Unadjusted overall survival between POUGIC subjects and controls

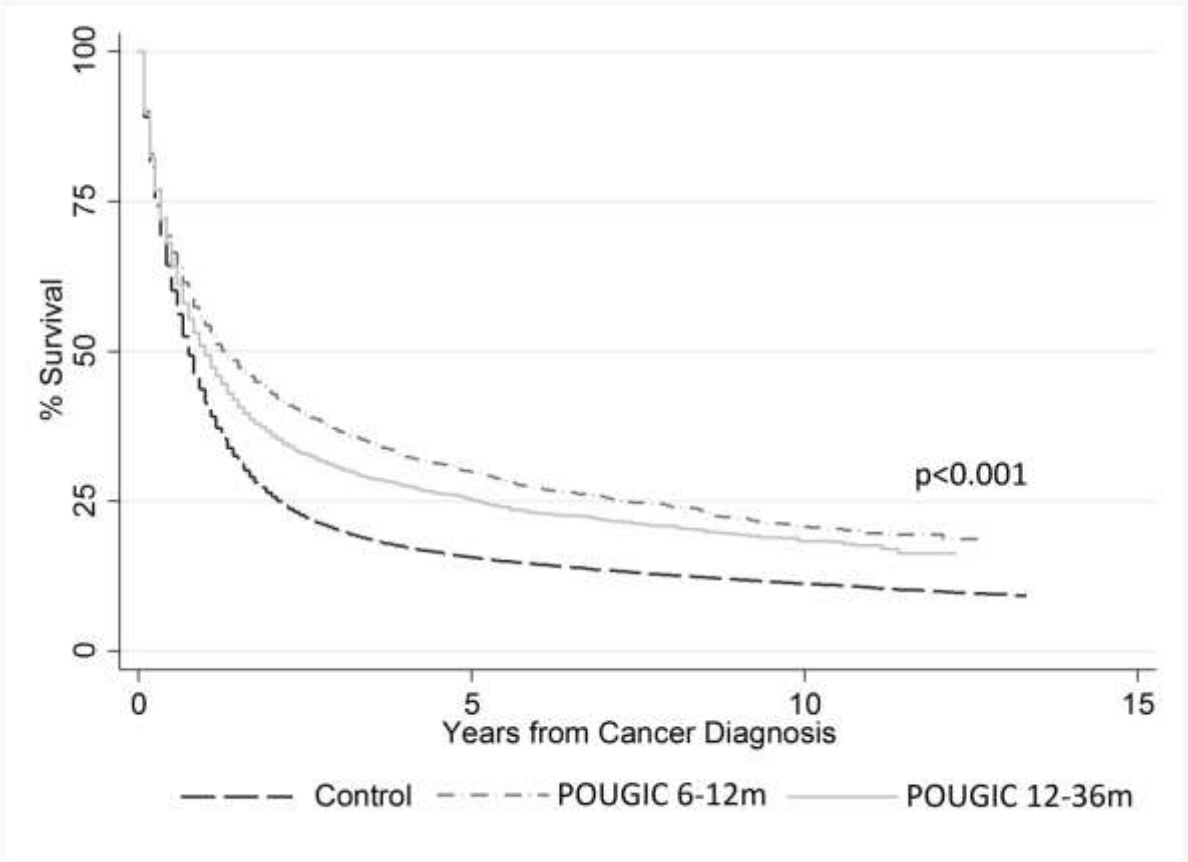


Figure 5.2 Unadjusted overall survival between POUGIC subjects with GC and controls

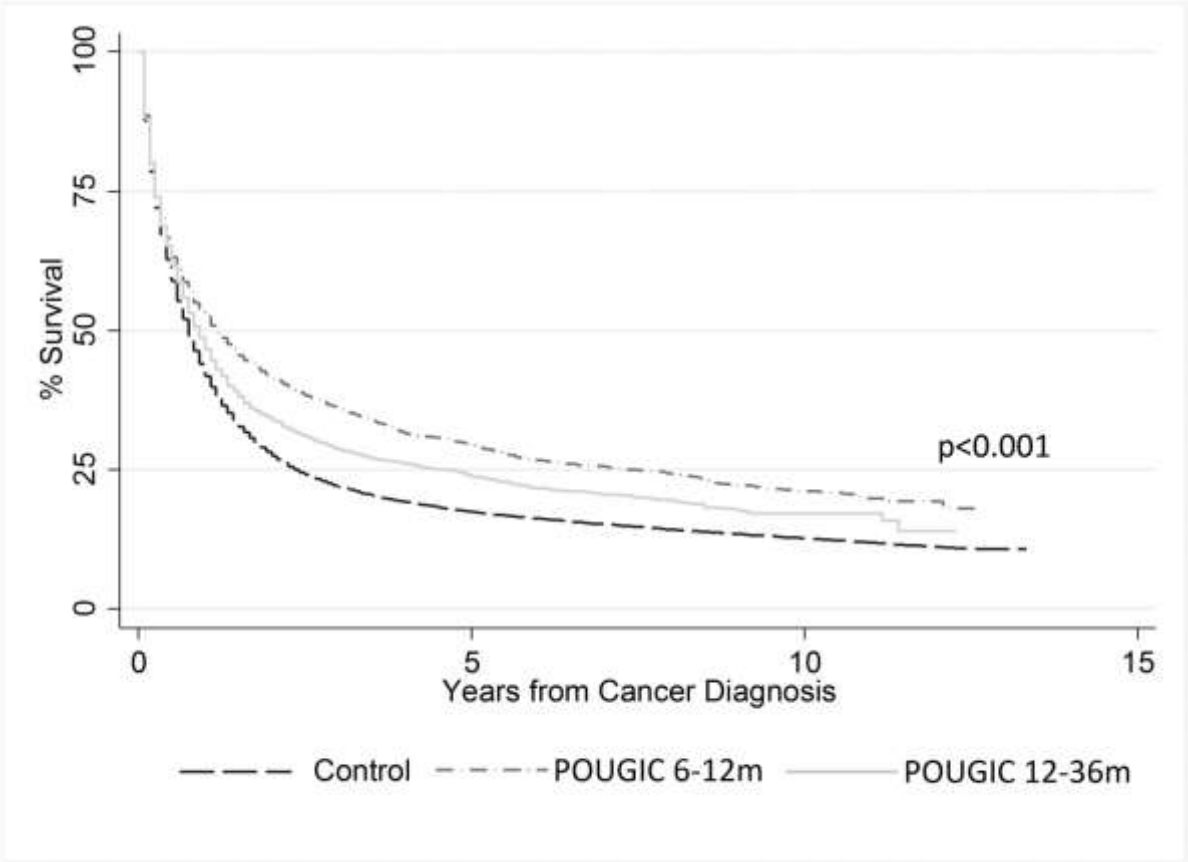
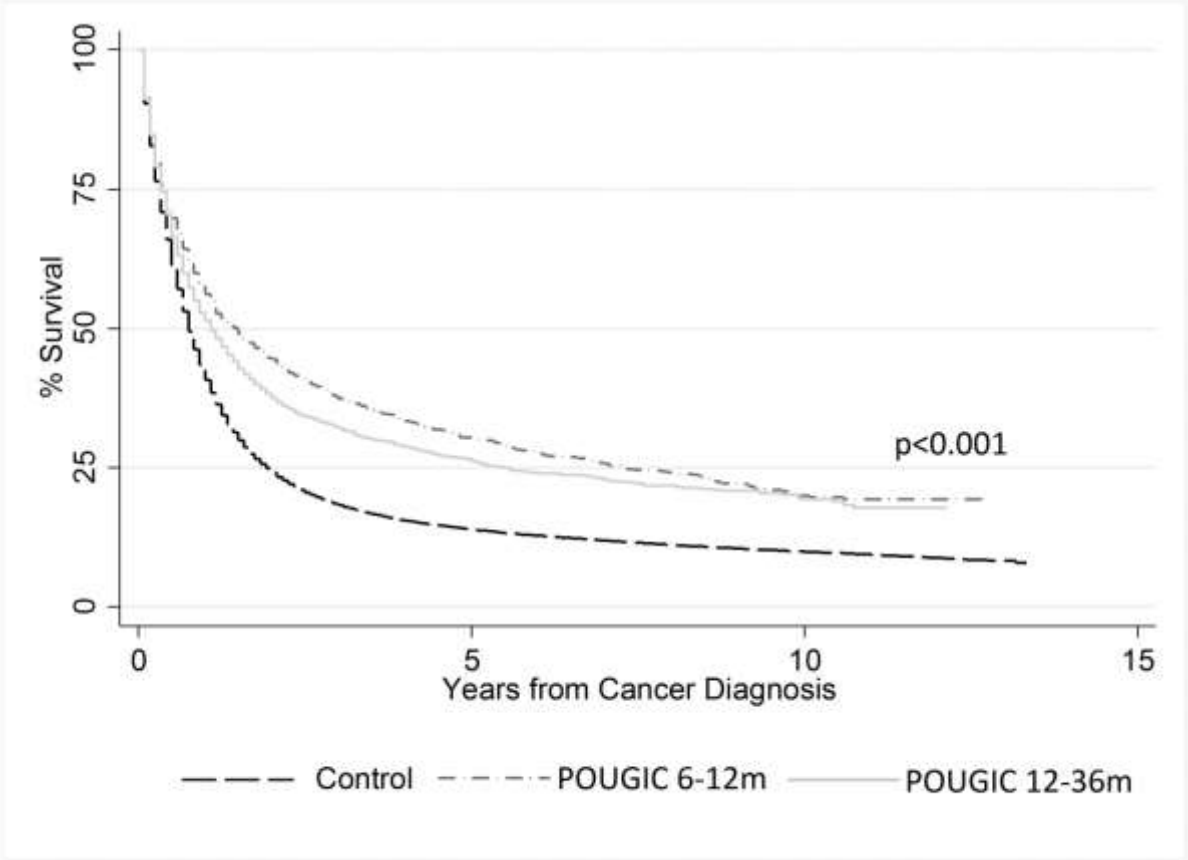


Figure 5.3 Unadjusted overall survival between POUGIC subjects with OC and controls



5.3.5 Service provider variables

The influence of provider variables on POUGIC are shown on Table 5.4. Medium volume providers (NHS trusts performed between 1092 to 3218 OGDs per year) where less likely to be associated with POUGIC (0.93 (95% CI 0.89-0.97), p=0.001) than high volume providers (NHS trust performed more than 3218 OGDs per year). Similarly, NHS providers with medium percentage (between 35% and 42%) of UGIC diagnosed during an emergency admission were less likely to be associated with POUGIC compared with NHS providers with low percentage of UGIC diagnosed (less than 35%) during an emergency admission (0.94 (95% CI 0.90-0.99), p=0.016). Gastric ulcer follow up performance appeared to be linked with POUGIC, with medium performing (24% to 39% of gastric ulcers were followed up) NHS

providers being protective of POUGIC (0.95 (95% CI 0.91-0.99), p=0.016) and low performing NHS providers (less than 24% of gastric ulcers were followed up) were more at risk of POUGIC (1.23 (95% CI 1.13-1.33), p<0.001) when compared with highest performing NHS providers (more than 39% of gastric ulcers were followed up). However, following adjusting for confounding factors, the POUGIC association with gastric ulcers follow up performance were no longer valid. PCCRC rates of NHS providers (see chapter 6) were investigated in relation with POUGIC, with NHS trusts separated into tertiles. Providers with lowest PCCRC rates (9.7% or less), providers with medium PCCRC rates (9.7% to 10.8%) and providers with highest PCCRC rates (over 10.8%) were compared. Once confounding factors were adjusted, providers with medium PCCRC rates were less likely to be associated with POUGIC (0.79 (95% CI 0.72-0.86), p<0.0001) but providers with highest PCCRC rates were more at risk of associated with POUGIC (1.43 (95% CI 1.31-1.56), p<0.001) when compared with providers with lowest PCCRC rates.

5.3.6 POUGIC rates over time

The number of OGDs performed, the number of UGICs and POUGIC rate by year is shown in Table 5.5. The number of OGD recorded in HES has increased by around 10% over the study period, however, the number of UGIC diagnosed per year remain static with around 12,000 UGIC cases diagnosed per year. The annual rate of POUGIC remains has remained broadly unchanged over the 10 year period. The POUGIC ranged from 10.0% to 12.4%.

5.4 Discussions

The overall POUGIC rate in England between 2003 and 2012 was 10.0% which is comparable with published POUGIC rates. The POUGIC rate reported by a meta-analysis of 3787 UGIC patients from 10 studies was 11.3%. As the meta-analysis included two Japanese studies with significantly higher PUOGIC rates of 25.8%[17] and 14.4%[19] respectively, which would explained the slightly higher overall POUGIC rate reported. Recent studies also using HES data have reported POUGIC rate in OC subjects of 7.8%[24] and 8.3%[67]. These reported figures appeared lower than the current study as only HES data from April 2011 to March 2012 were utilised in both of these studies, in addition, UGIC subjects have also been excluded if subjects were not captured in the National Oesophago-Gastric Cancer Audit, which has a case ascertainment rate of 83%.

POUGIC was associated with younger age as endoscopists may be less likely to expect malignancy in younger subjects and therefore less vigilant at examining the mucosa. Similarly, as UGIC is less prevalent in female subjects, this may also affect the attitude of the endoscopist during examination. Published studies also reported younger subjects[55, 68] and female subjects[68] have lower tolerance to upper GI endoscopy, which may contribute a poorer examination quality and less likely to detect an early and more subtle UGIC lesion. In HES, subjects with greater comorbidity were less likely to be associated with POUGIC, as previous study have reported that they are nearly 7 times more likely to have their UGIC diagnosed during an emergency admission, which is usually associated with more advanced disease.[42] Caucasian subjects are much more likely to be associated with OAC, in contrast, Asian or Afro-Caribbean subjects have a higher rate of OSCC. As OSCC is normally located at

upper oesophagus which is usually visualised less well than lower oesophagus where OAC is more prevalent. This may explain the increased POUGIC risk in Asian and Afro-Caribbean subjects.

Emergency OGDs are normally carried out for inpatients presenting with red flag symptoms (such as haemtemesis, melaena, vomiting) which are associated with advance UGIC therefore would be more likely to be detected at OGD. This would explain the negative association of emergency OGDs with POUGIC occurrence. This study also shown GC were more likely to be associated with POUGIC than OC subjects, in addition, OC subjects with distal oesophageal cancer were protective from POUGIC. As the incidence of OC was higher than GC during the study period with OAC accounting for two third of all OC, making lower oesophageal OC more likely to be detected when compared with early proximal oesophageal OC or early GC which are usually flatter (OSCC) or more subtle (atrophic gastritis) than more polypoid early OAC lesions. It was unsurprising that oesophageal ulcer was present in over one third and oesophageal stricture was reported one in every 15 POUGIC OGDs. Oesophageal ulcer can be associated Barrett's oesophagus which along with oesophageal strictures are associated with OC. Similarly, gastric ulcer was present in over a quarter of POUGIC OGD for subjects who eventually GC was detected at a later date. It was not possible to evaluate from HES if these subjects had adequate biopsies taken to exclude malignancy at their POUGIC OGD.

It was surprising liver metastases, peritoneal metastases and lymph node metastases were less likely in POUGIC subjects than controls. It is not clear if this is due screening bias in which POUGIC subjects having had a negative OGD prior to having a further examination soon after thus presenting at an earlier stage than control or if this also included subjects with Barrett's oesophagus that are on surveillance program with OC detected at earlier stage on screening. Barrett's oesophagus subjects with detected high grade dysplasia or early Barrett's cancer on surveillance OGD may be suitable for radiofrequency ablation or endoscopic mucosal resection. This may also explained the unadjusted survival advantage in POUGIC subjects when compared with controls and further supported by the effect most marked between POUGIC subjects with OC and OC controls.

Interestingly, there were no clear links between high OGD volume NHS providers and low OGD volume NHS providers or NHS providers with high percentage of UGIC diagnosed during an emergency admission than those with low percentage with POUGIC. Endoscopists performing OGDs tend to more heterogeneous group than other endoscopic procedures such as colonoscopy with nurse endoscopists, endoscopists in training and surgical endoscopists, which may explain the lack of association with procedure volume and POUGIC as some of these endoscpists may be of less experience and less likely to detect early and more subtle UGIC lesions. Gastric ulcer follow up performance appeared to be linked with POUGIC, with low performing providers more at risk of POUGIC but this effect was no longer valid following adjusting confounding factors which suggest the initial effect was due to subject selection. POUGIC may also be linked to general NHS provider endoscopy performance with providers with the highest PCCRC rate also more at risk of POUGIC. It was

unclear if this is due to endoscopist performance which may be training related or endoscopy departmental factors such as instrument quality.

Although the number of OGDs performed has increased slightly by around 10% over the study period, the number of UGIC diagnosed per year remained mostly unchanged over the same period. More significantly, the annual POUGIC rates were also static and ranged from 10.0% to 12.4% over the study period. In contrast to colonoscopy, the practice of OGD has not change significantly over the last decade with very limited key performance indicators (KPI) recommended by JAG.[66] In comparison, due to the spotlight on colonoscopy and enhanced KPI, the overall colonoscopy quality in the UK has improved substantially over the corresponding period.[30]

Using HES data has the benefit of examining POUGIC on a national level over a prolonged period. It allowed comparison between different chronological periods and also variation between providers. This study tackles all of these factors in which previously published studies were unable to address. Despite the obvious advantages with HES, there were some drawbacks and limitations. HES data has been validated in previous studies[24, 67, 69], however, as the data has to be manually entered either by clinicians or administrative staffs, concerns over data accuracy remains due to this step. As the HES data used in the study was unable to be linked with cancer registry records, one cannot assume all UGIC were captured despite validating processes as descripted above. Furthermore, the significant number of UGIC subjects with unspecified OC or GC location, lack of histological morphology, TNM

staging at diagnosis and if biopsies where taken at POUGIC ODGs were not available in HES and all would be useful in further evaluate the associations and potential causes of POUGIC.

Lastly, as ICD-10 code for Barrett's oesophagus (K22.7) was only routinely recorded in HES since 2008, this may over estimated the POUGIC rate as subjects on Barrett's surveillance program were not excluded from the overall POUGIC figure although the total number of subjects Barrett's oesophagus will be relatively small.

Table 5.1. The subject characteristics of post OGD upper gastrointestinal cancer cases and controls

	POUGIC 6-12 months	POUGIC 12-36 months	All POUGIC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)					**Univariate**			**Multivariate**		
	4029	7196	11225	101205						
Mean age (years±SD)										
	71.2±12.4	70.8±11.8	71.0±12.0	71.3±12.0			**0.003**			
Age group (number (%))										
Tertile 1 (youngest)	1200 (29.8%)	2160 (30.0%)	3360 (29.9%)	29737 (29.4%)	Ref	-	-	Ref	-	-
Tertile 2	1681 (41.7%)	3193 (44.4%)	4874 (43.4%)	44109 (43.6%)	0.98	0.93-1.02	0.346	1.01	0.95-1.08	0.672
Tertile 3 (oldest)	1148 (28.5%)	1843 (25.6%)	2991 (26.6%)	27359 (27.0%)	0.97	0.92-1.02	0.213	**0.88**	**0.81-0.95**	**0.001**
Gender (number (%))										
Male	2467 (61.2%)	4525 (62.9%)	6992 (62.3%)	66510 (65.7%)	Ref	-	-	Ref	-	-
Female	1562 (38.8%)	2671 (37.1%)	4233 (37.7%)	34695 (34.3%)	**1.16**	**1.11-1.21**	**<0.001**	**1.20**	**1.12-1.27**	**<0.001**
Charlson comorbidity index (number (%))										
0	3025 (75.1%)	5416 (75.3%)	8441 (75.2%)	70644 (69.8%)	Ref	-	-	Ref	-	-
1-4	383 (9.5%)	773 (10.7%)	1156 (10.3%)	9110 (9.0%)	1.06	0.99-1.14	0.071	1.08	0.98-1.19	0.106
5+	621 (15.4%)	1007 (14.0%)	1628 (14.5%)	21451 (21.2%)	**0.64**	**0.60-0.67**	**<0.001**	**0.80**	**0.74-1.19**	**<0.001**

Table 5.1 (cont.)	POUGIC 6-12 months	POUGIC 12-36 months	All POUGIC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)					**Univariate**			**Multivariate**		
	4029	7196	11225	101205						
Deprivation quintile (number (%))										
1 (most)	897 (22.3%)	1647 (22.9%)	2544 (22.7%)	21666 (21.4%)	Ref	-	-	Ref	-	-
2	834 (20.7%)	1517 (21.1%)	2351 (20.9%)	20899 (20.7%)	0.96	0.90-1.02	0.156	**0.91**	**0.83-1.00**	**0.046**
3	832 (20.7%)	1454 (20.2%)	2286 (20.4%)	21048 (20.8%)	**0.92**	**0.87-0.98**	**0.010**	0.93	0.85-1.02	0.133
4	776 (19.3%)	1362 (18.9%)	2138 (19.0%)	19902 (19.7%)	**0.91**	**0.86-0.97**	**0.004**	**0.89**	**0.82-0.98**	**0.017**
5 (least)	680 (16.9%)	1200 (16.7%)	1880 (16.7%)	17450 (17.2%)	**0.92**	**0.86-0.98**	**0.007**	**0.87**	**0.79-0.96**	**0.005**
Unknown	10 (0.2%)	16 (0.2%)	26 (0.2%)	240 (0.2%)	0.92	0.61-1.39	0.698			
Ethnicity (number (%))										
White or White British	3689 (91.6%)	6675 (92.8%)	10364 (92.3%)	91101 (90.0%)	Ref	-	-			
Asian or Asian British	67 (1.7%)	144 (2.0%)	211 (1.9%)	1354 (1.3%)	**1.37**	**1.18-1.59**	**0.001**			
Black or Black British	85 (2.1%)	127 (1.8%)	212 (1.9%)	1400 (1.4%)	**1.33**	**1.15-1.54**	**0.001**			
Chinese	8 (0.2%)	13 (0.2%)	21 (0.2%)	188 (0.2%)	0.98	0.63-1.54	0.937			
Mixed	7 (0.2%)	24 (0.3%)	31 (0.3%)	219 (0.2%)	1.24	0.85-1.81	0.255			
Other ethnic groups	27 (0.7%)	58 (0.8%)	85 (0.8%)	689 (0.7%)	1.08	0.86-1.36	0.483			
Unknown	146 (3.6%)	155 (2.2%)	301 (2.7%)	6254 (6.2%)	**0.42**	**0.38-0.48**	**0.001**			

Table 5.2. Upper GI endoscopy characteristics and findings of post OGD upper gastrointestinal cancer cases and controls

	POUGIC 6-12 months	POUGIC 12-36 months	All POUGIC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)					**Univariate**			**Multivariate**		
	4029	7196	11225	101205						
Procedure day (number (%))										
Weekday	3927 (97.5%)	6952 (96.6%)	10879 (96.9%)	98371 (97.2%)	Ref	-	-			
Weekend	102 (2.5%)	244 (3.4%)	346 (3.1%)	2834 (2.8%)	1.10	0.99-1.24	0.087			
Procedure type (number (%))										
Elective	3218 (79.9%)	5891 (81.9%)	9109 (81.1%)	73835 (73.0%)	Ref	-	-	Ref	-	-
Emergency	792 (19.7%)	1279 (17.8%)	2071 (18.4%)	27018 (26.7%)	**0.62**	**0.59-0.65**	**<0.001**	**0.77**	**0.71-0.84**	**<0.001**
Other	19 (0.5%)	26 (0.4%)	45 (0.4%)	352 (0.3%)	1.03	0.76-1.41	0.823	0.87	0.49-1.52	0.614
Upper gastrointestinal cancer location (number (%))										
Oesophagus	1976 (49.0%)	3813 (53.0%)	5789 (51.6%)	53352 (52.7%)	Ref	-	-			
Stomach	2053 (51.0%)	3383 (47.0%)	5436 (48.4%)	47853 (47.3%)	**1.05**	**1.01-1.09**	**0.021**			
Oesophageal cancer location (number (%))										
Proximal oesophagus	223 (11.3%)	565 (14.8%)	788 (13.6%)	8379 (15.7%)	Ref	-	-			
Distal oesophagus	593 (30.0%)	1375 (36.1%)	1968 (34.0%)	24048 (45.1%)	**0.87**	**0.80-0.95**	**0.002**			
Oesophagus NOS	1160 (58.7%)	1873 (49.1%)	3033 (52.4%)	20925 (39.2%)	**1.54**	**1.42-1.67**	**<0.001**			

Table 5.2 (cont.)	POUGIC 6-12 months	POUGIC 12-36 months	All POUGIC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)					**Univariate**			**Multivariate**		
	4029	7196	11225	101205						
Gastric cancer location (number (%))										
Proximal stomach	377 (18.4%)	782 (23.1%)	1159 (21.3%)	15032 (31.4%)	Ref	-	-			
Mid stomach	212 (10.3%)	403 (11.9%)	615 (11.3%)	7253 (15.2%)	1.10	0.99-1.22	0.067			
Distal stomach	210 (10.2%)	325 (9.6%)	535 (9.8%)	6034 (12.6%)	1.15	1.03-1.28	1.03			
Stomach NOS	1254 (61.1%)	1873 (55.4%)	3127 (57.5%)	19534 (40.8%)	**2.08**	**1.93-2.23**	**<0.001**			
Pathology detected in POUGIC procedures in subjects with OC (number (%))										
Oesophageal ulcer	760 (38.5%)	1470 (38.6%)	2230 (38.5%)							
Oesophageal stricture	202 (10.2%)	178 (4.7%)	380 (6.6%)							
Pathology detected in POUGIC procedures in subjects with GC (number (%))										
Gastric and/or peptic ulcer	755 (38.2%)	756 (19.8%)	1511 (26.1%)							

Table 5.3. The prevalence of metastases within 12 months of diagnosis and treatment outcomes in post OGD upper gastrointestinal cancer cases and controls

	POUGIC 6-12 months	POUGIC 12-36 months	POUGIC (all)	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)					**Univariate**			**Multivariate**		
	4029	7196	11225	101205						
Subjects with metastases within 12 months of diagnosis (number (%))										
Liver metastases	341 (8.5%)	844 (11.7%)	1185 (10.6%)	16190 (16.0%)	**0.62**	**0.58-0.66**	**<0.001**	**0.56**	**0.51-0.63**	**<0.001**
Lung metastases	217 (5.4%)	471 (6.5%)	688 (6.1%)	6892 (6.8%)	**0.89**	**0.82-0.97**	**0.006**	**0.84**	**0.79-0.89**	**<0.001**
Peritoneal metastases	208 (5.2%)	413 (5.7%)	621 (5.5%)	6239 (6.2%)	**0.89**	**0.82-0.97**	**0.008**	**0.75**	**0.59-0.95**	**0.017**
Bone metastases	42 (1.0%)	99 (1.4%)	141 (1.3%)	1324 (1.3%)	**0.81**	**0.72-0.90**	**<0.001**	0.87	0.75-1.01	0.072
Lymph node metastases	244 (6.1%)	585 (8.1%)	829 (7.4%)	12994 (12.8%)	**0.54**	**0.50-0.58**	**<0.001**	**0.55**	**0.49-0.61**	**<0.001**
Treatment outcome following diagnosis (number (%))										
Surgery	782 (19.4%)	1286 (17.9%)	2068 (18.4%)	21370 (21.1%)	**0.84**	**0.80-0.87**	**<0.001**			
Chemotherapy	785 (19.5%)	1795 (24.9%)	2580 (23.0%)	36801 (36.4%)	**0.52**	**0.50-0.55**	**<0.001**			
Radiotherapy	284 (7.0%)	596 (8.3%)	880 (7.8%)	8156 (8.1%)	0.97	0.90-1.04	0.418			

Table 5.4. The influence of provider variables on post OGD upper gastrointestinal cancer

	POUGIC 6-12 months	POUGIC 12-36 months	All POUGIC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)					**Univariate**			**Multivariate**		
	4029	7196	11225	101205						
Endoscopy unit volume by NHS provider (number (%))										
High volume providers (>3218/year)	2696 (66.9%)	4814 (66.9%)	7510 (66.9%)	66098 (65.3%)	Ref	-	-			
Medium volume providers (1092-3218/year)	1244 (30.9%)	2216 (30.8%)	3460 (30.8%)	32756 (32.4%)	**0.93**	**0.89-0.97**	**0.001**			
Low volume providers (<1092/year)	89 (2.2%)	166 (2.3%)	255 (2.3%)	2351 (2.3%)	0.95	0.84-1.09	0.489			
Percentage of UGIC diagnosed during an emergency admission by NHS provider (number (%))										
Low percentage providers (<35%)	1075 (26.7%)	1790 (24.9%)	2865 (25.5%)	25308 (25.0%)	Ref	-	-			
Medium percentage providers (35-42%)	1875 (46.5%)	3479 (48.3%)	5354 (47.7%)	50241 (49.6%)	**0.94**	**0.90-0.99**	**0.016**			
High percentage providers (>42%)	1079 (26.8%)	1927 (26.8%)	3006 (26.8%)	25656 (25.4%)	1.04	0.98-1.09	0.193			

Table 5.4 (cont.)	POUGIC 6-12 months	POUGIC 12-36 months	All POUGIC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)					**Univariate**			**Multivariate**		
	4029	7196	11225	101205						
Gastric ulcers follow up performance by NHS provider (number (%))										
Highest performing providers (>39%)	1709 (42.4%)	3014 (41.9%)	4723 (42.1%)	42014 (41.5%)	Ref	-	-	Ref	-	-
Medium performing providers (24-39%)	2047 (50.8%)	3694 (51.3%)	5741 (51.1%)	53681 (53.0%)	**0.95**	**0.91-0.99**	**0.016**	1.17	1.00-1.37	0.054
Low performing providers (< 24%)	273 (6.8%)	488 (6.8%)	761 (6.8%)	5509 (5.4%)	**1.23**	**1.13-1.33**	**<0.001**	0.95	0.87-1.03	0.232
Provider PCCRC rates										
Lowest (<9.7%)	981 (24.3%)	1577 (21.9%)	2558 (22.8%)	25073 (24.8%)	Ref	-	-	Ref	-	-
Medium (9.7-10.8%)	599 (14.9%)	1004 (14.0%)	1603 (14.3%)	15361 (15.2%)	1.02	0.96-1.09	0.499	**0.79**	**0.72-0.86**	**<0.001**
Highest (>10.8%)	1039 (25.8%)	1875 (26.1%)	2914 (26.0%)	26474 (26.2%)	**1.08**	**1.02-1.14**	**0.008**	**1.43**	**1.31-1.56**	**<0.001**
Not applicable	1410 (35.0%)	2740 (38.1%)	4150 (37.0%)	34297 (33.9%)	**1.19**	**1.13-1.24**	**<0.001**			

Table 5.5. Post OGD upper gastrointestinal cancer rates, total number of UGIC and OGDs by procedure year

	POUGIC 6-12 months	POUGIC 12-36 months	All POUGIC	Controls	Total UGIC	All gastroscopies	OR	95% CI	p value
2003-2004	433 (10.7%)	758 (10.5%)	1191 (10.6%)	11197 (11.1%)	12388	511486 (10.9%)	Ref	-	-
2004-2005	432 (10.7%)	682 (9.5%)	1114 (9.9%)	10639 (10.5%)	11753	492954 (10.5%)	0.98	0.90-1.07	0.720
2005-2006	413 (10.3%)	741 (10.3%)	1154 (10.3%)	10981 (10.9%)	12135	466627 (10.0%)	0.99	0.91-1.08	0.781
2006-2007	454 (11.3%)	756 (10.5%)	1210 (10.8%)	11303 (11.2%)	12513	479756 (10.2%)	1.01	0.93-1.09	0.881
2007-2008	415 (10.3%)	759 (10.5%)	1174 (10.5%)	11526 (11.4%)	12700	509585 (10.9%)	0.96	0.88-1.04	0.316
2008-2009	474 (11.8%)	779 (10.8%)	1253 (11.2%)	11466 (11.3%)	12719	535560 (11.4%)	1.03	0.95-1.12	0.526
2009-2010	484 (12.0%)	834 (11.6%)	1318 (11.7%)	11282 (11.1%)	12600	579629 (12.4%)	1.10	1.01-1.19	0.026
2010-2011	459 (11.4%)	863 (12.0%)	1322 (11.8%)	11473 (11.3%)	12795	545894 (11.7%)	1.09	1.00-1.18	0.057
2011-2012	465 (11.5%)	1024 (14.2%)	1489 (13.3%)	11338 (11.2%)	12827	560891 (12.0%)	**1.23**	**1.14-1.34**	**<0.001**
Total	4029 (100.0%)	7196 (100.0%)	11225 (100.0%)	101205 (100.0%)	112430	4682382 (100.0%)			

Chapter 6: Regional post OGD upper gastrointestinal cancer rate in the Midlands

6.1 Introduction

Although UGIC is less common than other types of cancer in the UK, the overall survival outcome of UGIC subjects remains poor compared with other types cancers and also with UGIC subjects in Western Europe.[44, 62] At present, the majority of UGIC subjects presents at advanced stages of disease, which limits their treatment option to palliative rather than curative intent,[70] this is likely to be the result of multiple factors. There are published reports suggesting that patients present late following developing UGIC symptoms [71] but also delay by primary care in referring these patients for diagnostic tests in secondary care.[72, 73] There are also growing evidence that subjects who had OGD which did not detect cancer but subsequently having UGIC diagnosed within 3 years, with these events termed POUGIC.[24, 26, 67] Published data on POUGIC to date are either with data from single provider units or with national hospital administrative data which both has its benefits and limitations. A meta-analysis include 10 studies published up to 2012 reported a POUGIC rate of 14%, as all the studies were from single provider units which invariably will be affected by subject selection bias and also subjected to endoscopists and institutional factors.[26] Subsequent studies from a same group investigated POUGIC rates in OC and GC diagnosed in 2011 to 2012 using a national hospital administration database.[24, 67] Although using a national database overcomes subject selection bias and institutional factors, the obvious disadvantage of inability to examine endoscopic factors such as endoscopist specialty, presence of trainee endoscopist, pre-medication usage, if

abnormalities were detected at POUGIC OGDs, biopsies taken and sample numbers which all maybe relevant to association of POUGIC.

6.2 Materials and methods

Electronic endoscopy records from 11 NHS trust in the Midlands were extracted from individual endoscopy databases following information governance approval. The endoscopy reports were then submitted to WMCIU for cancer registration linkage. Extra information submitted to WMCIU by individual NHS trust including type of UGIC, location of OC or GC, histology morphology, treatment outcomes and date of death were populated for matched UGIC subjects. UGIC subjects who had an OGD which did not detect UGIC between 90 days to 3 year before their eventual UGIC diagnosis were identified as POUGIC subjects with UGIC subjects only having OGD up to 90 days prior to their UGIC diagnosis act as controls. Subject characteristic, endoscopic factors and UGIC variables were extract from endoscopy reports and also from data provided from WMCIU. Logistic regression was used to identify and associated factors with UGIC occurrence.

6.3 Results

From the 11 NHS trust in the region, a total of 189,142 OGD records were sent to WMCIU for UGIC registration linkage. There were 6448 UGIC subjects identified, representing 3.4% of all OGDs submitted to WMCIU. There was a variation in UGIC detection rate between the NHS trusts in the Midlands ranging from 2.6% to 5.0%. Following excluding UGIC subjects who did not meet study criteria, a total of 4266 UGIC cases were analysed. Of these, 3817 (89.5%) subjects were classified as controls with 449 being POUGIC cases. Roughly half of POUGIC subjects had OGD 3-12months prior to UGIC diagnosis (224, 49.9%) and 225 (50.1%)

of POUGIC subjects had OGD 12-36 months prior to UGIC diagnosis. The overall POUGIC rate in the Midlands during the study period was 10.5%.

6.3.1 POUGIC and control characteristics

The characteristics of POUGIC subjects and controls are shown in Table 6.1. POUGIC subjects were younger the controls (70.0 years compared with 72.1 years, $p<0.0001$). There was no difference in gender and between subjects with OC or GC in the risk of association with POUGIC. When analysing OC subjects separately, there was no difference between subjects with OC at upper and mid oesophagus and those with OC at lower oesophagus with risk of POUGIC. Similarly, there was also no difference between subjects with OAC and OSCC with POUGIC occurrence. In GC subjects, subjects with GC at the body of stomach were more likely to be associated with POUGIC than subjects with GC at fundus and cardia. There was not difference between antrum and pyloric GC and fundus and cardia GC in relation to POUGIC risk. When analysing GC subjects by histology, there was no increase in POUGIC risk between GC subjects with diffused type adenocarcinoma and intestinal type adenocarcinoma.

Table 6.1. The characteristics of POUGIC subjects and UGIC controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
Number of subjects						
	449 (10.5%)	3817 (89.5%)	4266			
Age						
	70.0±11.4	72.1±11.5				**<0.0001**
Gender						
Male	295 (65.7%)	2528 (66.2%)	2823	Ref	-	-
Female	154 (34.3%)	1289 (33.8%)	1443	1.02	0.83-1.26	0.82
UGIC type						
OC	253 (56.3%)	2233 (58.5%)	2486	Ref	-	-
GC	196 (43.7%)	1584 (41.5%)	1780	1.09	0.90-1.33	0.38
OC location						
Upper and mid oesophagus	67 (26.5%)	667 (29.9%)	734	Ref	-	-
Lower oesophagus	147 (58.1%)	1359 (60.9%)	1506	1.08	0.80-1.46	0.63
Unknown/ overlapping sites	39 (15.4%)	207 (9.3%)	246	1.88	1.23-2.87	0.0037
OC histology						
Adenocarcinoma	134 (53.0%)	1376 (61.6%)	1510	Ref	-	-
Squamous cell carcinoma	75 (29.6%)	646 (28.9%)	721	1.19	0.89-1.61	0.25
Others	40 (15.8%)	201 (9.0%)	241	2.04	1.39-3.00	0.0003
Histology unknown	4 (1.6%)	10 (0.4%)	14	4.11	1.27-13.27	0.018
GC location						
Fundus and cardia	28 (14.3%)	379 (23.9%)	407	Ref	-	-
Body of stomach	60 (30.6%)	426 (26.9%)	486	**1.91**	**1.19-3.05**	**0.0071**
Antrum and pylorus	52 (26.5%)	469 (29.6%)	521	1.50	0.93-2.42	0.097
Unknown/ overlapping sites	56 (28.6%)	310 (19.6%)	366	2.45	1.52-3.94	0.0002
GC histology						
Diffused type Adenoca	11 (5.6%)	119 (7.5%)	130	Ref	-	-
Intestinal type Adenoca	33 (16.8%)	335 (21.1%)	368	1.07	0.52-2.18	0.86
Adenoca NOS	102 (52.0%)	910 (57.4%)	1012	1.21	0.63-2.32	0.56
Unknown/ others	50 (25.5%)	220 (13.9%)	270	2.46	1.23-4.90	0.01

6.3.2 Presenting symptoms of UGIC subjects

The presenting symptoms of UGIC subjects prior to OGD are shown in Table 6.2 and Table 6.3. For OC subjects, the presence of abdominal pain (1.97 (95% CI 1.21-3.22), p=0.0066), anaemia (1.78 (95% CI 1.14-2.77), p=0.01) and haematemesis and melanea (2.26 (95% CI 1.38-3.71), p=0.0013) all increased the risk of associated with POUGIC. In contrast, OC subjects presented with abnormal radiology (0.09 (95% CI 0.01-0.65), p=0.017), dysphagia (0.18 (95% CI 0.13-0.23), p<0.0001), weight loss (0.33 (95% CI 0.20-0.54), p<0.0001) or any red flag symptoms (0.20 (95% CI 0.15-0.26), p<0.0001) prior to undergoing OGD were less likely to be associated with POUGIC.

Table 6.2. Presenting symptoms in POUGIC subjects and controls with OC

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
Abdominal pain	21 (8.3%)	98 (4.4%)	119	**1.97**	**1.21-3.22**	**0.0066**
Abnormal radiology	1 (0.4%)	94 (4.2%)	95	**0.09**	**0.01-0.65**	**0.017**
Anaemia	26 (10.3%)	135 (6.0%)	161	**1.78**	**1.14-2.77**	**0.01**
Dysphagia	77 (30.4%)	1592 (71.3%)	1669	**0.18**	**0.13-0.23**	**<0.0001**
GORD	41 (16.2%)	284 (12.7%)	325	1.32	0.93-1.90	0.12
Haematemesis/ melaena	21 (8.3%)	86 (3.9%)	107	**2.26**	**1.38-3.71**	**0.0013**
Odynophagia	4 (1.6%)	68 (3.0%)	72	0.51	0.19-1.41	0.20
Vomiting	19 (7.5%)	151 (6.8%)	170	1.12	0.68-1.84	0.66
Weight loss	18 (7.1%)	421 (18.9%)	439	**0.33**	**0.20-0.54**	**<0.0001**
Red flag symptoms	150 (59.3%)	1965 (88.0%)	2115	**0.20**	**0.15-0.26**	**<0.0001**

For GC subjects, presenting with vomiting (0.37 (95% CI 0.20-0.67), p=0.001), weight loss (0.25 (95% CI 0.15-0.43), p<0.0001), dysphagia (0.63 (95% CI 0.41-0.97), p=0.036) and any red flag symptoms (0.38 (95% CI 0.28-0.52), p<0.0001) were protective of developing POUGIC. GC subjects who presented with GORD (1.71 (95% CI 1.22-2.40), p=0.0017) or

undergoing OGD for gastric ulcer healing check up (2.28 (95% CI 1.18-4.4), p=0.014) were around twice as likely to be associated with POUGIC.

Table 6.3. Presenting symptoms in POUGIC subjects and controls with GC

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
Abdominal pain	19 (9.7%)	194 (12.2%)	213	0.77	0.47-1.26	0.3000
Abnormal radiology	2 (1.0%)	50 (3.2%)	52	0.32	0.08-1.31	0.11
Anaemia	47 (24.0%)	395 (24.9%)	442	0.95	0.67-1.34	0.77
Dysphagia	26 (13.3%)	309 (19.5%)	345	**0.63**	**0.41-0.97**	**0.036**
GORD	55 (28.1%)	294 (18.6%)	349	**1.71**	**1.22-2.40**	**0.0017**
Haematemesis/ melaena	26 (13.3%)	201 (12.7%)	227	1.05	0.68-1.63	0.82
Odynophagia	1 (0.5%)	6 (0.4%)	7	1.35	0.16-11.3	0.78
Ulcer healing	12 (6.1%)	44 (2.8%)	56	**2.28**	**1.18-4.4**	**0.014**
Vomiting	12 (6.1%)	240 (15.2%)	252	**0.37**	**0.20-0.67**	**0.001**
Weight loss	16 (8.2%)	412 (26.0%)	428	**0.25**	**0.15-0.43**	**<0.0001**
Red flag symptoms	117 (59.7%)	1258 (79.4%)	1375	**0.38**	**0.28-0.52**	**<0.0001**

6.3.3 Endoscopic factors associated with POUGIC

The pre-medication use of OGD and sedation dosage are summarised on Table 6.4. UGIC subjects were separated into subjects with Pre-OGD medication of local anaesthesia (lidocaine throat spray), intravenous sedation (midazolam) and subjects received both local anaesthesia and intravenous sedation. UGIC subjects who were given midazolam were more at risk of developing POUGIC when compared with UGIC subjects who were not given midazolam prior to OGD (1.29 (95% CI 1.04-1.59), p=0.019). When sedation use was further investigated, older POUGIC subjects (age >70 years) had higher dosage of midazolam compared with control of the same age (3.4mg compared with 3.0mg, p=0.001). There was

no difference in midazolam dosage in younger (age <70 years) UGIC subjects between POUGIC subjects and controls.

Table 6.4. Pre-medication for OGD and sedation dosage

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
Pre-medication for OGD						
Spray	185(44.3%)	1692 (44.3%)	1877	0.88	0.72-1.07	0.21
Sedation	146 (32.5%)	1040 (27.2%)	1186	**1.29**	**1.04-1.59**	**0.019**
Both	85 (18.9%)	784 (20.5%)	869	0.90	0.70-1.16	0.42
Midazolam dose used						
Age (>70yrs)	3.4mg ±1.3	3.0mg ±1.3			**0.16-0.63**	**0.001**
Age (<70yrs)	4.1mg ± 1.4	3.9mg ±1.4			0.91-0.89	0.18

The endoscopist variables in POUGIC subjects and controls are shown on Table 6.5. When separating the lead endoscopist by specialty, there was no difference between medical endoscopists, surgical endoscopists and nurse endoscopists (and other specialties) with risk of developing POUGIC. In addition, the presence and involvement of a trainee (training endoscopist) was also not associated with POUGIC.

The finding at the OGDs which did not detect cancer (POUGIC OGDs) were examined, with 208 (46.3%) had abnormality reported at the same anatomical area in either the oesophagus or the stomach where UGIC was detected at a later date. Of the POUGIC OGDs, 115 (55.3%) were performed within 3-12 months of eventual UGIC diagnosis and 149 (71.6%) of POUGIC OGDs where abnormalities were reported had biopsies taken. For POUGIC subjects with OC, the most common pathology reported at POUGIC OGD was

oesophageal stricture in 24 OGDs (9.5%) and oesophagitis in 19 OGDs (7.5%). In POUGIC subjects with GC, the common pathology detected at POUGIC OGDs were gastritis (in 30 OGDs, 15.3%) and gastric ulcer (in 20 OGDs, 10.2%). POUGIC subjects had less number of biopsies taken (4.6 ±2.2) when compared with controls (5.4 ±2.2)(p=<0.0001).

Table 6.5. Endoscopist variables in POUGIC subjects and controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
Endoscopist specialty						
Medical	311 (69.3%)	2622 (68.7%)	2933	Ref	-	-
Surgical	72 (16.0%)	592 (15.5%)	664	1.03	0.78-1.35	0.86
Nursing/ other	26 (5.8%)	193 (5.1%)	219	1.14	0.74-1.74	0.67
Unknown	40 (8.9%)	410 (10.7%)	450	0.82	0.58-1.16	0.27
Trainee involvement						
No trainee	344 (76.6%)	2944 (77.1%)	3288	Ref	-	-
Trainee present	105 (23.4%)	873 (22.9%)	978	1.03	0.82-1.30	0.81

6.3.4 Treatment and survival outcomes for POUGIC subjects and controls

The treatment and survival outcomes for POUGIC subjects and controls are shown in Table 6.6. POUGIC subjects were more likely to undergo surgery (1.56 (95% CI 1.26-1.92), p<0.0001) but less likely chemotherapy (0.72 (95% CI 0.58-0.89), p=0.003) when compared with controls. There was no difference in radiotherapy between POUGIC subjects and controls. Surprisingly, POUGIC subjects had a survival advantage at 1 year post diagnosis (1.30 (95% CI 1.07-1.59), p=0.0086) when compared with controls.

Table 6.6. Treatment and survival outcomes for POUGIC subjects and controls

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
Treatment						
Surgery	154 (34.3%)	959 (25.1%)	1113	**1.56**	**1.26-1.92**	**<0.0001**
Chemotherapy	120 (26.7%)	1286 (33.7%)	1406	**0.72**	**0.58-0.89**	**0.003**
Radiotherapy	76 (16.9%)	670 (17.6%)	746	0.96	0.74-1.24	0.74
Survival						
Survival at 1yr post diagnosis	192 (42.8%)	1390 (36.4%)	1582	**1.30**	**1.07-1.59**	**0.0086**

6.3.5 POUGIC rate by providers and OGD date

There was a wide variation in POUGIC rate between individual NHS trusts in the Midlands (figure 6.1) with the rate ranged from 7.6% in the best performing trust to 12.4% in the worst performing trust. There was also no obvious link between the year in which OGDs were carried out (Table 6.7), except for examinations performed prior to 2000 being associated with POUGIC (1.79 (95% CI 1.29-2.49), p=0.0005). The change from fibre optic endoscopes to digital video endoscopes with superior image quality by most endoscopy units since late 1990s may explain the association with POUGIC in OGDs carried out pre-2000.

Figure 6.1. POUGIC rate by individual NHS trust

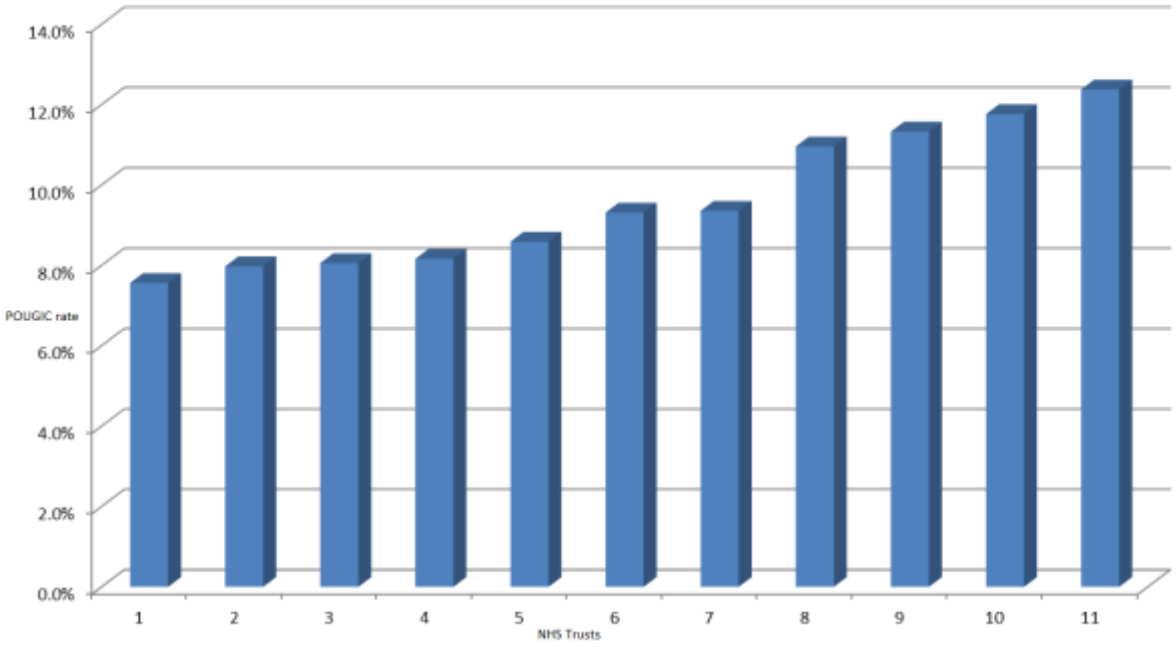


Table 6.7. POUGIC rates by OGD date

	POUGIC	Controls	Total	Odds ratio	95% CI	p value
Procedure date						
Pre 2000	51	265	316	**1.79**	**1.29-2.49**	**0.0005**
2000-2005	107	825	932	1.21	0.95-1.54	0.12
2005-2010	242	2255	2497	Ref	-	-
Post 2010	49	472	521	0.97	0.70-1.34	0.84
Total number of subjects	449	3817	4266			

6.4 Discussions

The overall POUGIC rate was 10.5% in the Midlands during the study period, although this ranged between 7.6% and 12.4% in the NHS trusts took part in the study. The rate at which UGIC was diagnosed also varied between 1 in every 20 to 40 OGDs but this was not correlated with the POUGIC rate.

POUGIC subjects were younger in the study, as younger subjects are less likely to present with red flag symptoms and as a result of this, endoscopists are less likely to expect malignancy than in the older subjects and may be less vigilant at examining the mucosa. In GC subjects, subjects with GC at the body of stomach were more likely to be associated with POUGIC as the incisura angularis is only adequately inspected on retroflexion, in addition, the posterior wall of the gastric body can often be difficult to visualise with can increase risk of failing to detect an early gastric lesion. Interestingly, there was no difference in POUGIC risk between GC subjects with diffused type adenocarcinoma and intestinal type adenocarcinoma, this may be due to the majority of GC subjects had no further histology differentiation recorded beyond adenocarcinoma which impaired statistical power in analysis.

It was unsurprising that OC subjects presented abnormal radiology, dysphagia, weight loss and any red flag symptoms and GC subjects presented with vomiting, weight loss, dysphagia and any red flag symptoms were all less likely to be associated with POUGIC as these symptoms are more common in advanced stages of UGIC and therefore more likely to be detected at OGD. Conversely, GC subjects presented with GORD are more likely to be associated with early stages of UGIC and OC subjects presented with haematemesis and melaena are likely have mucosal view impairment due to presence of blood in upper gastrointestinal tract and contribute to POUGIC. GC subjects undergoing OGD for gastric ulcer healing check up were over two folds more likely to be associated with POUGIC, this is expected given early GC can present as gastric ulcer which requires adequate number of biopsies and appropriate endoscopic follow up.

UGIC subjects who were given midazolam were more at risk of developing POUGIC when compared with UGIC subjects who were not given midazolam prior to OGD, in addition, older POUGIC subjects (age >70 years) had higher dosage of midazolam compared with control of the same age. Both findings were surprising, as one would expect patients under sedation would tolerate endoscopic examination better than those without. It may be the patient opted for sedation were more anxious[57, 74] or had unpleasant previous experience of OGD[68], resulting in this self selected cohort being less well tolerated during endoscopy examination and resulting in high risk of POUGIC occurrence. Reassuringly there was no different between medical endoscopists, surgical endoscopists and nurse endoscopists (and other specialties) with risk of developing POUGIC. In addition, the presence and involvement of a trainee (training endoscopist) was also not associated with POUGIC. This is unsurprising given performing a throughout OGD examination requires a systematic approach as the technical requirement is much less than other endoscopic procedures such as colonoscopy.[75]

Almost half of the POUGIC OGDs had abnormality reported at the same anatomical area in either the oesophagus or the stomach where UGIC was detected at a later date. Of these, over half were performed within 3-12 months of eventual UGIC diagnosis with the majority of these having biopsies taken at the time. POUGIC subjects had less number of biopsies taken when compared with controls which may contribute to POUGIC risk as previous studies have suggested taking up to 6 biopsies sample would be optimal in UGIC diagnosis.[27] Oesophageal strictures and gastric ulcers were both common finding on POUGIC OGDs which are expected as both are can be precursor to OC and GC respectively.

Surprisingly, POUGIC subjects were more likely to undergo surgery but less likely chemotherapy when compared with controls implies they would diagnosed at an earlier stage than controls. This may also explained the survival advantage at 1 year post diagnosis that POUGIC subjects had over controls. It is unclear as to the reason of this finding, with one possible explanation of this cohort having abnormality detected on POUGIC OGD but due to administrative delay and having repeat OGD after 90 days. This would classify these subjects as POUGIC but they will likely be diagnosed at an earlier stage of UGIC than controls.

There was no significant change in POUGIC rates over the study period which is unsurprising given there are very few national set standards for OGD and the practice remains unchanged over the last decade.

This current study is the largest and the only study using endoscopic data from multiple providers on POUGIC to date. The obvious advantages of using endoscopic records allowed the interrogation of variables and factors including endoscopist specialty, presence of trainee endoscopist, pre-medication usage, if abnormalities were detected at POUGIC OGDs, biopsies taken and sample numbers which all maybe relevant to the association of POUGIC with these information not recorded in national hospital administration database (such as HES). In sampling from multiple NHS providers across a region, this overcome subject selection bias and also institutional and endoscopist variation. In linking the endoscopic data with WMCIU cancer registration, it ensure the UGIC subject cohort is validated and robust and only subjects with a confirmed OC or GC were included in the study. Using endoscopic

records also posed a challenge as it was extremely time consuming and labour intensive as the data were recorded in no less than 7 different endoscopy reporting system. The endoscopy records relied on data input by the endoscopist, which one assume this is carried out accurately as it is impossible to independently validate the data. In addition, the data recorded at WMCIU can be incomplete, with a proportion of UGIC subjects without location or histological morphology of their UGIC.

Overall, the frequency of POUGIC in the Midlands remains relatively low and represents 1 POUGIC case in 420 OGDs. However, given the poor survival of POUGIC subjects, further research into factors contributing to POUGIC need to be carried out to ensure UGIC are diagnosed at the earliest opportunity.

Chapter 7: Factors associated with colorectal cancer occurrence after colonoscopy that did not diagnose colorectal cancer (an analysis of English Hospital Episode Statistics from 2003 to 2012).

(Data from this chapter has been published: **Cheung D, Evison F, Patel P, Trudgill N. Factors associated with colorectal cancer occurrence after colonoscopy that did not diagnose colorectal cancer. Gastrointestinal endoscopy 2016;84:287-95 e1.**)

7.1 Introduction

Colonoscopy is the gold standard for diagnosing, screening and surveillance for CRC. In England, the setting of national standards for colonoscopy and accreditation of endoscopy units has resulted in an improvement in auditable colonoscopy standards over the last decade.[30] The same period has also coincided with an increase in 5 year survival following CRC diagnosis from 47.8% to 53.6%.[76] However, 2.6 to 6.0% of CRC patients have previously been reported to be diagnosed within 5 years of a colonoscopy which did not detect cancer. These events are termed PCCRC.[28, 33, 77] It has been proposed that PCCRC may have a different cell biology from other CRC with more aggressive and rapidly growing tumours.[78, 79] However, two recently published North American studies concluded that this did not apply to the majority of PCCRC, with around two thirds of PCCRC a result of missed lesions or incomplete polypectomy.[77, 80]

Given the improvements in colonoscopy over the past decade in England, we have examined the impact on PCCRC in a national hospital episode database and associated risk factors for these events.

7.2 Materials and Methods

In order to prevent provider variation and patient selection bias, the frequency of PCCRC at national level and associated risk factors for these events were examined using a national hospital episode database in England. HES data between 2003 and 2012 were used for this study.

7.3 Results

7.3.1 Study cohort

Between April 2003 and March 2009, 1,439,684 colonoscopies were identified and 67,202 subjects were diagnosed with CRC during this period. Out of the 67,202 CRC subjects, there were 8,147 (12.1%) PCCRC subjects: 1796 (2.7%) PCCRC 6-12 months; 3,772 (5.6%) PCCRC 12-36 months and 2,579 (3.8%) PCCRC 36-60 months. 59,055 CRC subjects had not had a colonoscopy between 6 and 60 months prior to CRC diagnosis and served as controls. Overall, 0.66% or 1 in every 150 subjects developed PCCRC after a colonoscopy that did not diagnose CRC.

7.3.2 Validation of colonoscopy and colorectal cancer populations

The total number of colonoscopies carried out between 2007 and 2010 at UHB was 8708 and 8292 colonoscopies (95.2%) were coded in HES for UHB for the equivalent four year period. The CRC population was validated by comparing CRC cases recorded in HES (315,515) to CRC cases reported from 2002 to 2011 by NCIN (312,984)[44], showing a concordance of over 99%. The CRC population was further validated by comparing the 70.4% surgical rate for CRC from HES with the National Bowel Cancer Audit, which reported that 75.7% of CRC patients enrolled in the audit underwent surgery between 2008 and 2011.[45, 46, 47] All of the validation processes showed a good correlation between HES data and independent data sources, suggesting the study methodology was valid.

7.3.3 Subject characteristics

The characteristics of cases with PCCRC and CRC controls are shown in Table 7.1. PCCRC subjects (mean age 70.7±11.4 years) were older than controls (mean age 70.2±11.4 years)($p<0.001$). The risk of PCCRC appeared to increase with age on univariate analysis, but only subjects aged 70 to 74 were associated with PCCRC compared with subjects under 60, following adjusting for confounding factors. PCCRC subjects were more likely to be female. Subjects with the most co-morbidities (Charlson co-morbidity index of 5 or greater) were associated with PCCRC. PCCRC was not associated with differences in ethnicity or deprivation.

7.3.4 Colonoscopy variables and findings

The influence of colonoscopy variables and findings on PCCRC are shown in Table 7.2. The majority of CRC were diagnosed during an elective colonoscopy. However, being diagnosed during an emergency colonoscopy reduced the risk of PCCRC nearly by half. There was minor increased risk of PCCRC on univariate analysis in colonoscopies carried out at the weekend compared with during the week.

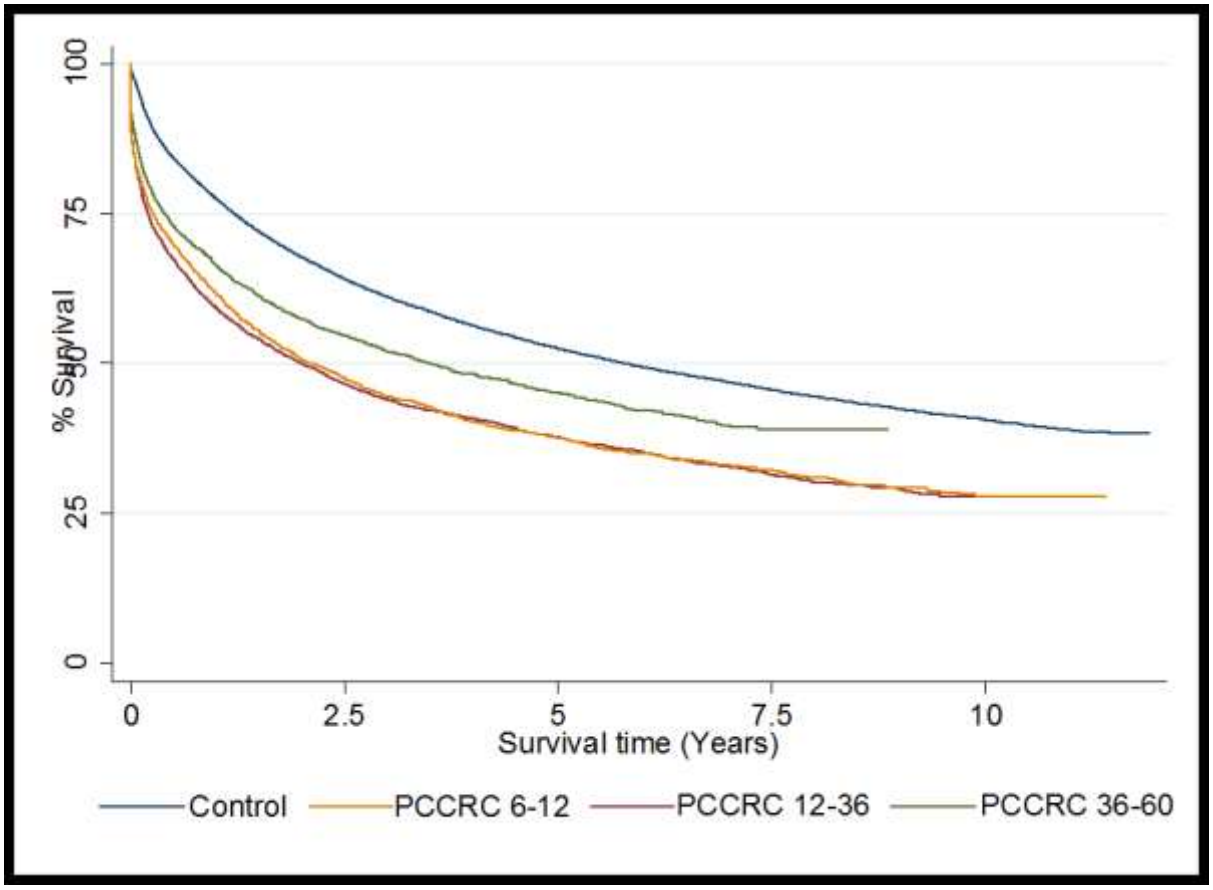
PCCRC was more likely to be associated with CRC in the right colon. Colonic polyps were coded in 21.6% of the colonoscopies which did not detect CRC in the PCCRC group. Polypectomy was coded in a further 18.9%. On univariate analysis, this was higher than both the recorded polyp rate of 9.8% (2.52 (95% CI 2.39-2.65), p<0.0001) and polypectomy rate of 11.3% (1.82 (95% CI 1.72-1.92), p<0.0001) from all colonoscopies during the study period. Furthermore, the polyp and polypectomy rates were both higher in the PCCRC 6-12 months group on univariate analysis, than in the PCCRC 12-36 months (p<0.0001) and PCCRC 36-60 months (p<0.0001) groups.

7.3.5 Colorectal outcomes and survival

The prevalence of metastatic disease within 12 months of CRC diagnosis in PCCRC cases and controls are shown in Table 7.3. PCCRC cases were up to twice as likely to be diagnosed with lung, peritoneal and bone metastases within 12 months of CRC diagnosis. However, lymph node metastases were more prevalent in controls than PCCRC cases, suggesting coding bias related to the increased rate of surgery in control subjects described later.

On univariate analysis, PCCRC cases were less likely to undergo surgery compared with controls (0.33 (95% CI 0.32-0.35), p<0.0001) or chemotherapy (0.66 (95% CI 0.62-0.69), p<0.0001). Overall survival was also worse in PCCRC subjects compared with controls, with a median survival of 5.8 years in controls compared with 2.1 years in the PCCRC 6-12 months group, 2.0 years in the PCCRC 12-36 months group and 3.5 years in the PCCRC 36-60 months group (figure 6.1). Following adjusting for age, gender, co-morbidity and deprivation, survival outcomes remained worse for PCCRC subjects with a hazard ratio of 1.17 (95% CI 1.10- 1.24)(p<0.0001), 1.26 (95% CI 1.20-1.31)(p<0.0001) and 1.20 (95% CI 1.13-1.27)(p<0.0001) for the PCCRC 6-12 months, PCCRC 12-36 months and PCCRC 36-60 months respectively when compared with controls.

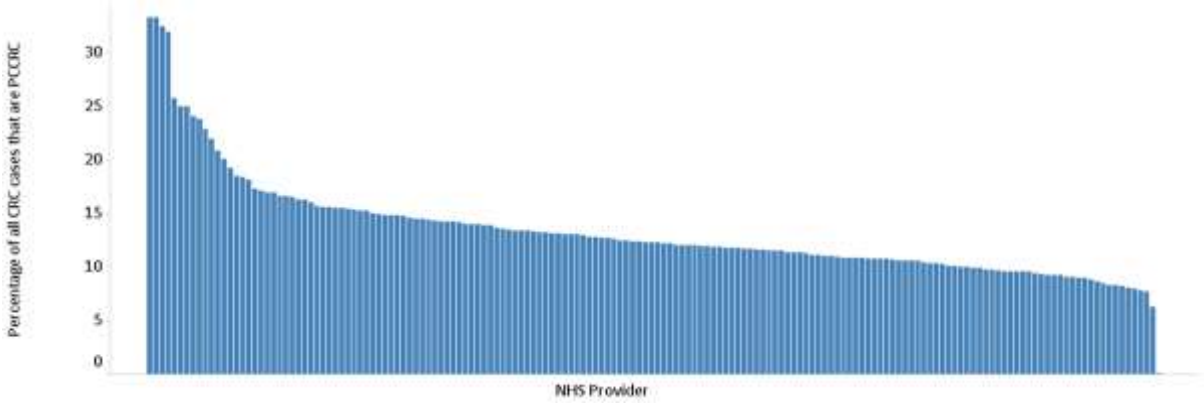
Figure 7.1. Unadjusted survival following colorectal cancer diagnosis in post-colonoscopy colorectal cancer cases and control subjects



7.3.6 Individual provider variables

The influence of provider variables on PCCRC are shown in Table 7.4. There was a more than twofold variation in PCCRC rates between individual providers in England during the study period (figure 7.2). On univariate analysis, medium colonoscopy volume providers and low volume providers were both more likely to be associated with PCCRC than high volume providers. Following adjusting for other variables in the multivariate model an association with medium volume providers remained. BCSP accreditation status and the percentage of CRC diagnosed as an emergency were not associated with an increased risk of PCCRC.

Figure 7.2. Post-colonoscopy colorectal cancer rates by individual provider in England between 2003 and 2009

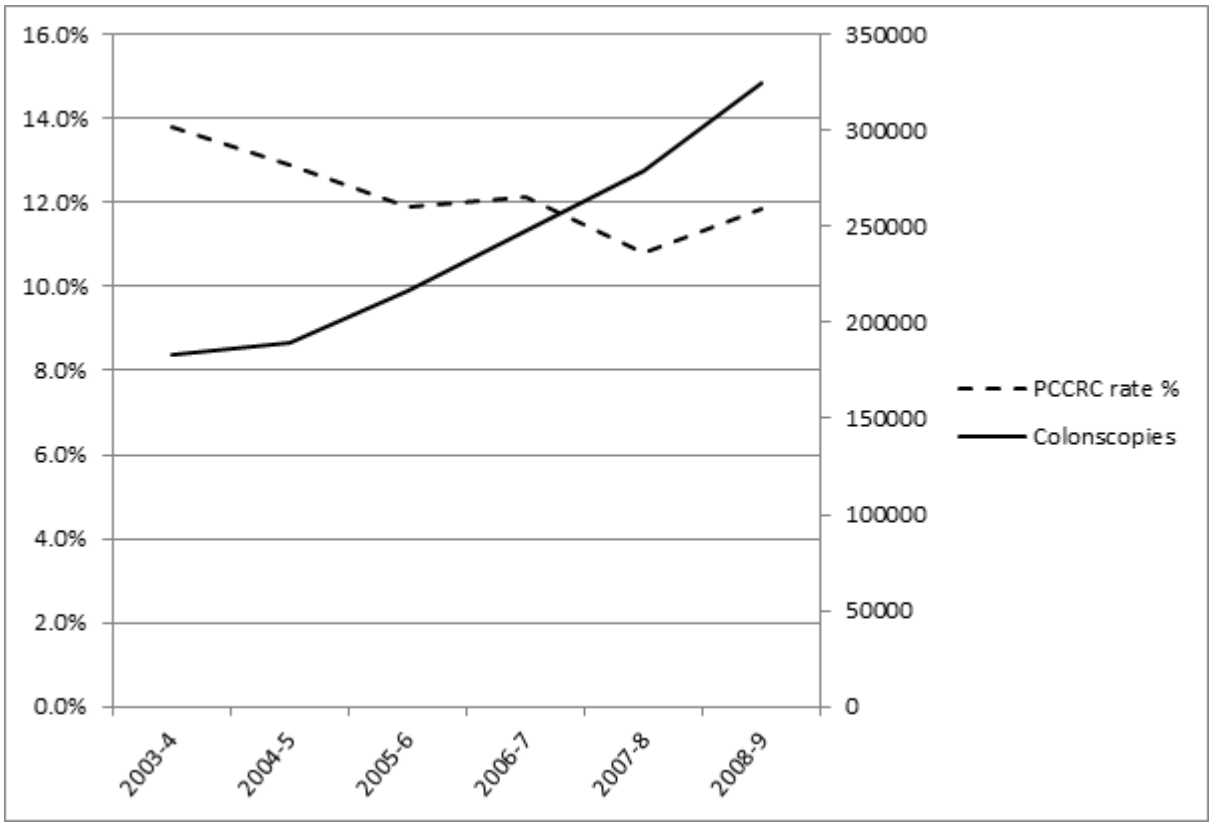


7.3.7 PCCRC rates over time

The number of colonoscopies recorded in HES has increased by almost two fold over the study period. Despite the increase in colonoscopy numbers performed year on year, the annual rate of PCCRC has steadily fallen over the study period ($p<0.0001$)(figure 7.3). The annual PCCRC rate decreased from 13.8% in 2003-2004 to 11.9% by the end of study period

in 2008-2009 with the reduction seen mainly in the PCCRC 6-12 months and PCCRC 12-36 months groups.

Figure 7.3 Post-colonoscopy colorectal cancer rates and colonoscopy volume in England by year



7.4 Discussion

The overall PCCRC rate of 12.1% in 67202 subjects in England between 2003 and 2009 appears higher than previously published figures. However, some previous studies have calculated the PCCRC rate by only including CRC subjects with a colonoscopy up to 36 months prior to diagnosis and the comparable figure from the present study is 8.3%. A Canadian study of 14,064 CRC subjects reported a PCCRC rate of 9.0% between 2000 and

2005.[35] Using the Surveillance, Epidemiology, and End Results (SEER)-Medicare database in the USA, a PCCRC rate of 7.2% was reported between 1994 to 2005 from a study of 57,839 CRC subjects.[36] A further population based study from Utah, USA with 2659 CRC subjects between 1995 and 2009 described a PCCRC rate of 6% when subjects with a colonoscopy up to 60 months prior to CRC diagnosis were included.[77] In Europe, two recent studies have reported much lower PCCRC rates. A Danish population based study between 2000 to 2009 included 37,044 CRC subjects and concluded that only 2.7% of CRC subjects have had a colonoscopy that failed to diagnose CRC 1 to 5 years prior to diagnosis.[33] A second study from the Netherlands analysed 5107 CRC subjects between 2001 to 2010 from three providers and found a PCCRC rate of only 2.9% for subjects with a colonoscopy up to 60 months prior to CRC diagnosis.[81] In addition to potential variations in subject and colonoscopy factors between the difference studies, the wide range of reported PCCRC rates are likely to be contributed to by methodological differences.[31]

In the present study, PCCRC was associated with older subjects, female gender, an increased number of co-morbidities and right-sided CRC, which is in keeping with findings from other studies of PCCRC. [28, 34, 35, 36] The association between increasing age and PCCRC was less marked on multivariate analysis and this may relate to confounding from increasing co-morbidity in the elderly. Elderly patients are more likely to have inadequate bowel preparation, thus reducing mucosal visualisation and detection of polyps and early CRC.[82, 83] Female patients are more likely to have had previous abdominal and pelvic surgery, which may increase the technical difficulty of colonoscopy and impair patient tolerance, reducing the caecal intubation rate.[84] In addition to factors that have an adverse effect on caecal intubation rate, right sided CRC are more likely to arise from flat, non-polypoid

adenomatous lesions[81, 85] that poor bowel preparation may make difficult to detect. This will contribute to the association of right sided CRC with PCCRC.

Over a fifth of PCCRC subjects had colonic polyps or polypectomy coded during the most recent colonoscopy prior to CRC diagnosis. This is higher than the average polypectomy rate in all colonoscopy procedures during the same period. Furthermore, polyp and polypectomy coding rates were highest in the PCCRC 12-36 months group. Prior polypectomy has been reported to double the risk of PCCRC[36], with up to 19% of CRC occurring in the same anatomic segment as a previously resected adenoma.[80] Paradoxically, colonoscopists with higher polypectomy rates have been reported to be associated with a lower risk of PCCRC[35, 36], presumably as they detect more polyps and remove them more completely than other colonoscopists. Incomplete polypectomy, or inadequate biopsy sampling of polyps, is therefore a key modifiable risk factor for PCCRC and ensuring adequate follow up and assessment following polypectomy may reduce PCCRC rates.

PCCRC subjects appeared to have worse outcomes in terms of both treatment following diagnosis (surgery and chemotherapy) and overall survival. Previous studies have reported no survival difference between PCCRC subjects and controls[33, 34] with one recent study even reporting a survival benefit in the PCCRC subjects, which was likely to be due to earlier CRC stage at diagnosis in the PCCRC subjects.[77] In the current study, PCCRC subjects were older, had greater co-morbidities and were more likely to present with distant metastases within 12 months of diagnosis compared with controls. All these factors contributed to the reduced rates of curative surgery or palliative chemotherapy for PCCRC subjects and will have contributed to worse survival. Adjusting the survival analyses for differences in ages,

gender, co-morbidity and deprivation still revealed worse survival for PCCRC subjects and, at least in England, PCCRC is clearly associated with worse survival. Survival in PCCRC subjects would have been potentially better if earlier opportunities to diagnose their CRC had been taken.

Previous studies have reported that PCCRC was not associated with endoscopist procedure volume[35] and that higher colonoscopy volumes may even be positively associated with PCCRC surprisingly.[36] In the current study, there was a large variation in PCCRC rates between individual providers across England but PCCRC appeared to be associated with lower colonoscopy volume providers.

Colonoscopy undertaken during an emergency admission covered 10% of procedures examined and was associated with a lower risk of PCCRC at 9% compared with 14% for elective procedures. Patients presenting as an emergency may have more advanced colorectal cancer and therefore a lower chance of PCCRC.

The annual PCCRC rate in England has fallen steadily over the study period from 13.8% to 11.9%, at least partly due to improving colonoscopy standards over the corresponding time period. In 2003, a multi-regional audit in England including 9223 colonoscopies reported that mean caecal intubation rate was only 76.9%.[86] A subsequent national audit in 2011 of 20085 colonoscopies found that the caecal intubation rate had improved to 92.3%.[30]

The use of a national hospital dataset enabled us to undertake one of the largest PCCRC studies to date, including the vast majority of colonoscopies performed during a period of

rising colonoscopy standards. The quality of diagnostic and procedural coding in HES has been previously investigated and there was a high concordance when compared with independent national data sources.[30, 42, 87] However, we did not have the opportunity to link our HES dataset directly to cancer registry data due to restrictions under which the data is held and therefore, in order to validate the methodology chosen, colonoscopy and CRC populations were compared with national cancer databases and a local data sample and revealed a good correlation. The completeness and accuracy of coding in HES is still a potential source of concern. For example, the diagnosis date may not be recorded accurately in HES due to the need for histological confirmation before CRC coding and therefore a colonoscopy within 6 months of CRC coding had to be considered the diagnostic procedure. There are also limitations in HES concerning coding of colonoscopy procedures, polyps, polypectomy, presence of metastases, surgery and chemotherapy and the figures included may be an over or under estimate, though this is likely to affect PCCRC cases and controls equally. A further limitation is that key procedure information such as the bowel preparation quality, sedation doses, colonoscopist grade and specialty, extent of examination, completeness of polypectomy and number of biopsies taken are not recorded in HES and all may influence the PCCRC risk. Furthermore, due to the HES coding hierarchy, indication, presence of diverticular disease and history of abdominal or pelvic surgery may not be coded, partly due to under reporting by colonoscopists when significant pathology or CRC are found and again each may be important risk factors for PCCRC. As HES does not record polyp histology or the International Classification of Diseases for Oncology (ICD-O) codes, the lack of data on polyp and CRC histology and Duke's staging further limits analysis of potential causes of PCCRC (de novo CRC, incomplete adenoma resection, missed lesion or biopsy failed to detect CRC) and survival in PCCRC subjects.

In conclusion, the PCCRC rate was 12.1% in England between 2003 and 2009. PCCRC was associated with older age, female gender, increasing co-morbidity, procedure related factors (elective procedures and right sided CRC) and provider colonoscopy volume.

Table 7.1. The characteristics of post-colonoscopy colorectal cancer cases and controls

	PCCRC 6-12 months	PCCRC 12-36 months	PCCRC 36-60 months	All PCCRC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)						**Univariate**			**Multivariate**		
	1796 (2.7)	3772 (5.6)	2579 (3.8)	8147 (12.1)	59055 (87.9)						
Mean age ±SD (years)											
	71.5±11.4	70.9±11.7	69.8±10.8	70.7±11.4	70.2±11.4			**<0.001**			
Age group (number (%))											
Under 60	263 (3.2)	598 (7.3)	415 (5.1)	1276 (15.7)	9849 (16.7)	Ref					
60-64	167 (2.0)	367 (4.5)	288 (3.5)	822 (10.1)	6749 (11.4)	0.94	0.86-1.03	0.1928	0.95	0.86-1.04	0.277
65-69	217 (2.7)	531 (6.5)	435 (5.3)	1183 (14.5)	8810 (14.9)	1.04	0.95-1.13	0.4044	1.03	0.94-1.12	0.537
70-74	344 (4.2)	648 (8.0)	488 (6.0)	1480 (18.2)	10229 (17.3)	**1.12**	**1.03-1.21**	**0.0067**	**1.09**	**1.00-1.18**	**0.039**
75-79	359 (4.4)	678 (8.3)	499 (6.1)	1536 (18.9)	10698 (18.1)	**1.11**	**1.02-1.20**	**0.0109**	1.07	0.98-1.16	0.159
80+	446 (5.5)	950 (11.7)	454 (5.6)	1850 (22.7)	12720 (21.5)	**1.12**	**1.04-1.21**	**0.0029**	1.08	1.00-1.17	0.065
Gender (number (%))											
Male	974 (12.0)	1974 (24.2)	1340 (16.4)	4288 (52.6)	33057 (56.0)	Ref	-	-	Ref	-	-
Female	822 (10.1)	1798 (22.1)	1239 (15.2)	3859 (47.4)	25998 (44.0)	**1.14**	**1.09-1.20**	**<0.0001**	**1.13**	**1.08-1.19**	**<0.001**
Charlson co-morbidity index (number (%))											
0	1514 (18.6)	3210 (39.4)	2235 (27.4)	6959 (85.4)	50663 (85.8)	Ref	-	-	Ref	-	-
1-4	154 (1.9)	298 (3.7)	210 (2.6)	662 (8.1)	4764 (8.1)	1.01	0.93-1.10	0.7896	1.06	0.97-1.16	0.195
5+	128 (1.6)	264 (3.2)	134 (1.6)	526 (6.5)	3628 (6.1)	1.06	0.96-1.16	0.2641	**1.16**	**1.05-1.28**	**0.003**

Table 7.1 (cont.)	PCCRC 6-12 months	PCCRC 12-36 months	PCCRC 36-60 months	All PCCRC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)						**Univariate**			**Multivariate**		
	1796 (2.7)	3772 (5.6)	2579 (3.8)	8147 (12.1)	59055 (87.9)						
Deprivation quintile (number (%))											
1 (most)	329 (4.0)	637 (7.8)	393 (4.8)	1359 (16.7)	10015 (17.0)	Ref	-	-			
2	365 (4.5)	740 (9.1)	499 (6.1)	1604 (19.7)	11258 (19.1)	1.05	0.97-1.13	0.2153			
3	333 (4.1)	782 (9.6)	551 (6.8)	1666 (20.4)	12399 (21.0)	0.99	0.91-1.07	0.8002			
4	387 (4.8)	784 (9.6)	568 (7.0)	1739 (21.3)	12642 (21.4)	1.01	0.94-1.09	0.7242			
5 (least)	381 (4.7)	823 (10.1)	566 (6.9)	1770 (21.7)	12620 (21.4)	1.03	0.96-1.11	0.3905			
Ethnicity (number (%))											
Caucasian	1656 (20.3)	3536 (43.4)	2467 (30.3)	7659 (94.0)	54512 (92.3)	Ref	-	-			
Asian	21 (0.3)	55 (0.7)	36 (0.4)	112 (1.4)	788 (1.3)	1.01	0.83-1.23	0.9097			
Afro-Caribbean	25 (0.3)	53 (0.7)	27 (0.3)	105 (1.3)	823 (1.4)	0.91	0.74-1.11	0.3553			
Chinese	0	0	0	12 (0.1)	118 (0.2)	0.72	0.40-1.30	0.2865			
Mixed	0	0	0	18 (0.2)	160 (0.3)	0.80	0.49-1.30	0.3719			
Others	12 (0.1)	21 (0.3)	21 (0.3)	54 (0.7)	341 (0.6)	1.13	0.85-1.50	0.4156			
Unknown	74 (0.9)	95 (1.2)	18 (0.2)	187 (2.3)	2313 (3.9)	**0.58**	**0.49-0.67**	**<0.0001**			

Table 7.2. The colonoscopy characteristics and findings of post-colonoscopy colorectal cancer cases and controls

	PCCRC 6-12 months	PCCRC 12-36 months	PCCRC 36-60 months	All PCCRC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)						**Univariate**			**Multivariate**		
	1796 (2.7)	3772 (5.6)	2579 (3.8)	8147 (12.1)	59055 (87.9)						
Procedure day (number (%))											
Weekday	1736 (21.3)	3628 (44.5)	2486 (30.5)	7850 (96.4)	57249 (96.9)	Ref	-	-			
Weekend	60 (0.7)	144 (1.8)	93 (1.1)	297 (3.6)	1806 (3.1)	**1.19**	**1.06-1.36**	**0.0044**			
Procedure type (number (%))											
Elective	1622 (19.9)	3473 (42.6)	2455 (30.1)	7550 (92.7)	52605 (89.1)	Ref	-	-	Ref	-	-
Emergency	174 (2.1)	299 (3.7)	124 (1.5)	597 (7.3)	6450 (10.9)	**0.64**	**0.59-0.70**	**<0.0001**	**0.54**	**0.59-0.69**	**<0.0001**
Colorectal cancer location (number (%))											
Left sided	897 (11.0)	1754 (21.5)	1260 (15.5)	3911 (48.0)	34703 (58.8)	Ref	-	-	Ref	-	-
Right sided	535 (6.6)	1242 (15.2)	919 (11.3)	2696 (33.1)	20751 (35.1)	**1.15**	**1.09-1.21**	**<0.0001**	**1.17**	**1.11-1.23**	**<0.0001**
Unknown/ overlapping sites	364 (4.5)	776 (9.5)	400 (4.9)	1540 (18.9)	3601 (6.1)	**3.79**	**3.54-4.06**	**<0.0001**	**3.72**	**3.46-3.99**	**<0.0001**

Table 7.2 (cont.)	PCCRC 6-12 months	PCCRC 12-36 months	PCCRC 36-60 months	All PCCRC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)						**Univariate**			**Multivariate**		
	1796 (2.7)	3772 (5.6)	2579 (3.8)	8147 (12.1)	59055 (87.9)						
Polyp/ polypectomy coded (number (%))											
Polyp coded	491 (6.0)	742 (9.1)	523 (6.4)	1756 (21.6)	141799* (9.8)	**2.52+**	**2.39-2.65+**	**<0.0001+**			
No polyp coded	1305 (16.0)	3030 (37.2)	2056 (25.2)	6391 (78.4)	1300714* (90.2)	Ref	-	-			
Polypectomy coded	348 (4.3)	669 (8.2)	523 (6.4)	1540 (18.9)	162364* (11.3)	**1.82+**	**1.72-1.92+**	**<0.0001+**			
No polypectomy coded	1448 (17.8)	3103 (38.1)	2056 (25.2)	6607 (81.1)	1280150* (89.7)	Ref	-	-			

*** From all colonoscopies, + Univariate analysis comparing all PCCRC with all colonoscopies during study period.**

Table 7.3. The prevalence of metastases within 12 months of colorectal cancer diagnosis in post-colonoscopy colorectal cancer cases and controls

	PCCRC 6-12 months	PCCRC 12-36 months	PCCRC 36-60 months	All PCCRC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)						**Univariate**			**Multivariate**		
	1796 (2.7)	3772 (5.6)	2579 (3.8)	8147 (12.1)	59055 (87.9)						
Subjects with metastases within 12 months of diagnosis (number (%))											
Liver metastases	276 (3.4)	619 (7.6)	365 (4.5)	1260 (15.5)	8545 (14.5)	**1.08**	**1.01-1.15**	**0.017**	0.97	0.91-1.05	0.486
Lung metastases	154 (1.9)	345 (4.2)	182 (2.2)	681 (8.4)	3104 (5.3)	**1.64**	**1.51-1.79**	**<0.0001**	**1.61**	**1.46-1.77**	**<0.0001**
Peritoneal metastases	75 (0.9)	166 (2.0)	102 (1.3)	343 (4.2)	1903 (3.2)	**1.32**	**1.17-1.48**	**<0.0001**	**1.27**	**1.12-1.44**	**<0.0001**
Bone metastases	45 (0.6)	106 (1.3)	78 (1.0)	229 (2.8)	678 (1.1)	**2.49**	**2.14-2.90**	**<0.0001**	**2.21**	**1.88-2.60**	**<0.0001**
Lymph node metastases	136 (1.7)	282 (3.5)	231 (2.8)	649 (8.0)	6459 (10.9)	**0.70**	**0.65-0.76**	**<0.0001**	**0.75**	**0.69-0.82**	**<0.0001**
Treatment outcome following diagnosis (number (%))											
Surgery	791 (9.7)	1661 (20.4)	1337 (16.4)	3789 (46.5)	42790 (72.5)	**0.33**	**0.32-0.35**	**<0.0001**			
Chemotherapy	422 (5.2)	911 (11.2)	594 (7.3)	1927 (23.7)	18908 (32.0)	**0.66**	**0.62-0.69**	**<0.0001**			

Table 7.4. The influence of provider variables on post-colonoscopy colorectal cancer

	PCCRC 6-12 months	PCCRC 12-36 months	PCCRC 36-60 months	All PCCRC	Controls	Odds ratio	95% CI	p value	Odds ratio	95% CI	p value
Total subjects (number)						**Univariate**			**Multivariate**		
	1796 (2.7)	3772 (5.6)	2579 (3.8)	8147 (12.1)	59055 (87.9)						
Colonoscopy volume by NHS provider (number (%))											
High volume providers (>1680 pa)	955 (11.7)	1993 (24.5)	1415 (17.4)	4363 (53.6)	33353 (56.5)	Ref	-	-	Ref	-	-
Medium volume providers	704 (8.6)	1486 (18.2)	994 (12.2)	3184 (39.1)	21942 (37.2)	**1.11**	**1.06-1.16**	**<0.0001**	**1.13**	**1.01-1.27**	**0.035**
Low volume providers (<747 pa)	137 (1.7)	293 (3.6)	170 (2.1)	600 (7.4)	3760 (6.4)	**1.22**	**1.11-1.34**	**<0.0001**	1.05	0.98-1.12	0.161
BCSP status (number (%))											
BCSP provider	959 (11.8)	2064 (25.3)	1396 (17.1)	4419 (54.2)	31780 (53.8)	Ref	-	-	-	-	-
Non-BCSP provider	837 (10.3)	1708 (21.0)	1183 (14.5)	3728 (45.8)	27275 (46.2)	0.98	0.94-1.03	0.4690	0.96	0.90-1.03	0.255
Percentage of CRC diagnosed during an emergency admission by NHS provider (number (%))											
Low percentage providers (<27.3%)	408 (5.0)	848 (10.4)	629 (7.7)	1885 (23.1)	14270 (24.2)	**0.91**	**0.84-0.98**	**0.0115**	0.96	0.87- 1.06	0.443
Medium percentage providers	1068 (13.1)	2273 (27.9)	1530 (18.8)	4871 (59.8)	35211 (59.6)	0.95	0.89-1.01	0.1299	0.96	0.85-1.09	0.531
High percentage providers (>33.9%)	320 (3.9)	651 (8.0)	420 (5.2)	1391 (17.1)	9572 (16.2)	Ref	-	-	Ref	-	-

Chapter 8: Summary and conclusions

This body of work investigated the frequency and association of POUGIC and PCCRC at a national and at a regional level. The studies contained in this body of work included the most number of UGIC subjects in a POUGIC study to date in the UK and Worldwide. By using endoscopic records from multiple providers, this was the first study to date for investigating POUGIC. The PCRRC study using HES data was the largest and covering the longest period in the UK to date. All the studies have provided unique data regarding POUGIC and PCCRC in the UK.

8.1 UGIC and POUGIC

OGD is the gold standard in visualising and diagnosing pathology in the upper gastrointestinal tract. Despite being the most frequently performed endoscopic procedure in the UK, there are very few quality standards set for OGD at present. The current quality markers for OGD set out by JAG consist only of successful intubation and completion of procedure (in reaching the second part of duodenum). In contrast, the standard in colonoscopy have improved dramatically over the last decade[30] since measurable indicators such as completion rate (caecal intubation rate and rectal retroversion rate), time spent on examining mucosal surfaces (colonoscopy withdrawal time), pathology detection rate (adenoma and polyp detection rate) and mucosa visualisation (bowel preparation quality) were introduced by JAG.[66]

The key findings of the POUGIC studies in this body of work has been summarised in table 8.1. The studies in this work showed despite OGD is a reliable test, POUGIC, the failing to

diagnosed OC or GC in within 36 months of eventual diagnosis in UGIC subjects is relatively common. The overall POUGIC rate was broadly similar between the above studies using different data source. The overall POUGIC rate was 6.7% from THIN (chapter 4), 10.0% from HES (chapter 5) and 10.5% from endoscopic data and WMCIU (chapter 6). These figures are comparable to the POUGIC rates reported from smaller UK based studies.[12, 13, 24, 67] The discrepancy in the POUGIC rates between the studies reflects slight variation in methodology and POUGIC definition used due to different data source. The POUGIC definition in the THIN study was subjects who underwent a negative OGD between 12 to 36 months prior to eventual UGIC diagnoses, whereas the duration of this period was 6 to 36 months in the HES study and 3 to 36 months in the WMCIU study respectively. The difference in POUGIC rate may also reflect the data were from different geographical areas with THIN data from a representation of primary care centres in the UK, HES covering secondary care (NHS Hospitals) in England and WMCIU data was linkage with endoscopy records from NHS hospitals in the Midlands.

Younger subjects and female gender were both associated factors for POUGIC despite UGIC are less common in these subject groups. As previously discussed, younger subjects [54, 55, 56] and female subjects [55, 57] reportedly having a lower tolerance for OGD examination, which may in turn lead to a reduction in OGD diagnostic quality. Poor subject tolerance leads to intra-procedural discomfort and may lead to early termination of the endoscopy and reduce diagnostic quality of the OGD. A recent study reported longer examination time during OGD were twice as likely to detect high risk lesions and three times more likely to detect GC irrespective of endoscopist experience.[88] In addition, as UGIC is less prevalent in these groups, there may be lower expectation by endoscopists to detect cancer. Asian

and Afro-Caribbean subjects were more at risk of developing POUGIC which may be due to OSCC are more prevalence in this population, and unlike the readily recognisable signs of early OAC such as Barrett's oesophagus, the early signs of OSCC may be less readily recognised by endoscopists.[14, 59] Subjects with GC and particularly GC at the body of stomach were associated with increased risk of POUGIC, this may be due to endoscopist technique as the incisura angularis is only adequately inspected on retroflexion, in addition, the posterior wall of the gastric body can often be difficult to visualise with can increase risk of failing to detect an early gastric lesion.

Symptoms on presentation were also associated with POUGIC with subjects presented with red flag (alarm) symptoms much less likely to be associated with POUGIC than subjects who present with non-red flag symptoms such as GORD. As red flag symptoms are associated with more advanced stages of UGIC, which are more likely to be detected on OGD examination. OC subjects presented with haematemesis and melaena and GC subjects undergoing OGD for gastric ulcer healing follow up were also more likely to associate with POUGIC occurrence. In subjects with haematemesis or melaena, mucosal visibility may be reduced by blood in the upper GI tract and subjects diagnosed with GU should have follow up OGD to check for ulcer healing, failure to arrange appropriate OGD follow up in these subject groups may explain the their association with POUGIC. Subjects requesting midazolam for OGD and those requiring a higher dose of midazolam were more likely to be associated with POUGIC as they may be more anxious and tolerate OGD less well than subjects who opted for local anaesthesia.

In order to reduce the POUGIC incidence, a high quality index OGD needs to be carried out. Both patient and endoscopic factors can affect the quality of an OGD. A high quality OGD should comprise of a comfortable and settled patient with an endoscopic view of the upper gastrointestinal tract free of mucus and debris to enable a thorough examination to be performed. Patient comfort and tolerance level to OGD affects endoscopic view more directly than colonoscopy. Patient factors such as younger age, female gender, high BMI, poor view of oropharyngeal anatomy and high apprehension and anxiety prior to OGD are reported to be associated with a lower tolerance for examination.[54, 55, 56, 57, 68, 89, 90, 91] Poor subject tolerance leads to intra-procedural discomfort and reduces diagnostic quality. Endoscopists should be aware of these factors and consider if sedation would be the pre-procedure medication of choice to ensure an adequate endoscopic examination in a comfortable patient. If a suboptimal OGD occurs, the endoscopist should have a low threshold for repeating the procedure to ensure pathology is excluded. In addition, trans-nasal OGD should also be offered to patients who are known to have a low tolerance of OGD as there are good evidence suggest it is better tolerated than traditional tran-oral OGD.[92, 93]

The visual quality of the upper gastrointestinal tract can be impaired by the presence of air bubbles, foam and mucus adhering the mucosa surface. The need for aspiration and use of lavages to clear the mucosa leads to longer examination time and is not carried out universally. Oral anti-foaming agents such as dimethicone and simethicone (Gascon) and mucolytics such as Pronase have been reported to improve endoscopic mucosal visibility of the upper gastrointestinal tract and pathology detection rate when used prior to or during OGD[94, 95, 96]. Despite the minimal additional time required by the patient and the

negligible associated extra cost, the use of these agents has not been taken up as standard practice in the UK to date.

As OGD is an observer dependent procedure, capturing of still images is the standard practice in Japan where up to 40 images are taken for a diagnostic procedure.[97] Although taking such number of images is unlikely to be practical or necessary in the Western population given the lower incidence of UGIC and early GC. The European Society of Gastrointestinal Endoscopy (ESGE) recommended recording still images in 8 different areas of the upper gastrointestinal tract during a diagnostic OGD over a decade ago,[75] this practice remained poorly adopted in the UK with most endoscopists only recording still images when pathology is detected. Adopting such practice would ensure an adequate quality OGD is carried out as it would require a relatively comfortable patient and good mucosal visibility for the images to be captured.

Almost half of the POUGIC OGDs had abnormality reported at the same anatomical area in either the oesophagus or the stomach where UGIC was detected at a later date. Endoscopic findings of oesophageal ulcer, oesophageal stricture and gastric ulcer were commonly reported at POUGIC OGD and also associated with POUGIC occurrence. Significantly, only 50.0% of subjects with oesophageal stricture or ulcer and 64.6% of subjects with gastric ulcer had a follow up OGD within 90 days from the THIN data study. The WMCIU study also demonstrated that POUGIC subjects had less number of biopsies taken when compared with controls which may contribute to POUGIC risk as previous studies have suggested taking up to 6 biopsies sample would be optimal in UGIC diagnosis.[27] A lack of adequate follow up of these lesions may also be a contributing factor to POUGIC cases.

This body of work has also demonstrated the risk of POUGIC varies between providers. Using WMCIU and endoscopy records, the POUGIC rate varies in the Midlands was shown to be 7.6% in the best performing provider to the highest of 12.4%. On a national level, the providers with the highest PCCRC rates were also more at risk of POUGIC occurrence. In order to address the provider variability, national quality standard would need to revised by the Joint Assessment Group (JAG) as the current KPI consist of successful intubation and completion of procedure (in reaching the second part of duodenum) is probably inadequate. The revised KPI should include patient comfort score, mucosal view, standardised imaging to document throughout and complete examination as suggested by ESGE.

Although there were conflicting evidence from the studies carried out in this body of work, with POUGIC subjects less likely to have liver metastases, peritoneal metastases and lymph node metastases than in controls, more likely to undergo surgery and even unadjusted survival advantage in POUGIC subjects, the overall survival outcome in all UGIC subjects were extremely poor. The survival at 1 year following UGIC diagnosis were only between 30% to 40%.

The POUGIC rates were also static from the data of the larger cohort (HES) which is of concern. There was evidence that older OGD were more likely to associate with POUGIC in the THIN and WMCIU studies, but this may be explained by replacement of fibre optic endoscope with superior image quality digital endoscope. The POUGIC rate would likely remains unchanged unless intervention on a national level. Further research into OGD quality and POUGIC should be encouraged by JAG and setting quality standards for OGD

that are similar to the stringent quality standards for colonoscopy that have improved outcomes for colonoscopy and colorectal cancer over the last decade. [30] POUGIC rates should also be regularly audited by endoscopy units and undertake root cause analysis of identified POUGIC cases. Providers should also submit POUGIC data to JAG as part of their annual assessment.

Table 8.1 Summary of the POUGIC studies and associations with POUGIC

	THIN (chapter 4)	HES (chapter 5)	WMCIU (chapter 6)
Population	UK	England	Midlands
Data source	Primary care database	Secondary care administrative database	Secondary care endoscopic database
POUGIC definition (OGD prior to UGIC diagnosis)	12-36 months	6-36 months *(inclusive of subjects with Barrett's)*	3-36 months
UGIC subjects	9487	112430	6448
POUGIC rate	**6.7%**	**10.0%**	**10.5%**
Associations with POUGIC			
Age	Younger age‡	Younger age*	Younger age*
Gender	Female‡	Female‡	No diff*
GC/OC	OG‡	GC*	No diff*
Comorbidity	Increasing comorbidity‡	Increasing comorbidity‡	n/a
Deprivation	Increased deprivation‡	Increased deprivation‡	n/a
Ethnicity	n/a	Asian, black*	n/a
Alarm symptoms[+]	Negative association*	n/a	Negative association*

OC location	No diff*	Proximal OC*	No diff*
GC location	No diff*	Mid stomach (vs proximal stomach)*	Body of stomach (vs fundus/cardia)*
Findings at POUGIC OGD prior to UGIC diagnosis			
OC subjects	Oesophageal stricture/ ulcer 5.7% Oesophagitis 19.4%	Oesophageal ulcer 38.5% Oesophageal stricture 6.6%	Oesophageal stricture 9.5% Oesophagitis 7.5%
GC subjects	Gastric ulcer 10.5% Gastritis 22.6%	Gastric ulcer 26.1%	Gastric ulcer 10.2% Gastritis 15.3%
Treatments and outcomes for POUGIC subjects (odds ratio)			
Surgery	No diff‡	0.84*	1.56*
Chemotherapy	No diff*	0.52*	0.72*
Radiotherapy	1.38‡	No diff*	No diff*
Survival at 1 year post UGIC diagnosis	no diff‡	no diff‡	1.30*
Time trend			
Time period when OGD performed	Association with earlier years*	No diff*	Association with earlier years *

*Univariate analysis

‡Multivariate analysis

+At diagnostic OGD for control subjects and POUGIC OGD for POUGIC subjects

8.2 CRC and PCCRC

Despite the improvement in colonoscopy standards and colorectal cancer outcome over the last decade,[30] PCCRC remains a relatively common in CRC subjects. Using HES data, the PCCRC rate was 12.1% in England between 2003 and 2009.

PCCRC was associated with older subjects, female gender and an increased number of co-morbidities. All these factors are likely to affect bowel preparation quality and caecal intubation rate. Elderly patients are more likely to have inadequate bowel preparation, with female patients are more likely to have had previous abdominal and pelvic surgery, both may increase the technical difficulty of colonoscopy and impair patient tolerance, reducing the caecal intubation rate.[84] Incomplete colonoscopy and reduced mucosal visualisation leading to failure of detection of polyps and early CRC and increase risk of PCCRC.[82, 83] Right sided CRC were shown to be associated with PCCRC, as right sided CRC are more likely to arise from flat, non-polypoid adenomatous lesions[81, 85] that poor bowel preparation may make difficult to detect. In view of these findings, subject factors should be taken into consideration when requesting a colonoscopy and alternative investigation methods (such a CT colonography) should be considered in patient where bowel preparation outcome is likely to be suboptimal and caecal intubation may be difficult. In addition, there is currently no requirement to document successful caecal intubation and introducing mandatory imaging capture of anatomical landmarks (appendiceal orifice, tri-radial fold, ileo-caecal valve or ileal mucosa) would ensure caecal intubation is achieved.

Over a fifth of PCCRC subjects had colonic polyps or polypectomy coded during the most recent colonoscopy prior to CRC diagnosis with polyp and polypectomy coding rates were highest in the subjects having had colonoscopy between 12 to 36 months prior to eventual CRC diagnosis. Prior polypectomy is associated with increased risk of PCCRC[36, 80], this is likely related to colonoscopist experience and incomplete polypectomy.[35, 36] In order to address this, individual providers should ensure adequate follow up and assessment following polypectomy but also to consider reducing the number of colonoscopists carrying out more difficult polypectomies may reduce PCCRC rates.

There was a large variation in PCCRC rates between individual providers across England but PCCRC appeared to be associated with lower colonoscopy volume providers. This result should be further investigated and correlated with colonoscopy quality indicators such as caecal intubation, adenoma detection rates and perhaps colonoscopist procedure volume that are likely to be potentially more important factors in PCCRC incidence.

PCCRC subjects had worse outcomes both in treatment following diagnosis and overall survival. Survival in PCCRC subjects may have been potentially better if earlier opportunities to diagnose their CRC had been taken. The improving PCCRC rate is likely to continue to improve in recent years given changes in colonoscopy practice, including the recognition of the importance of minimum withdrawal times[98], bowel preparation improvements[99] and better endoscopic recognition of sessile serrated polyps[85], subsequent to the study period. PCCRC rate should be a routinely measured endoscopy unit colonoscopy quality marker and potentially avoidable risk factors for PCCRC addressed.

Chapter 9: Publications and presentations

All or part of the following chapters were presented at Gastroenterological meeting and published as original articles or abstracts in peer reviewed journals.

How common is upper gastrointestinal cancer occurrence after endoscopy that did not diagnose cancer (A UK primary care based study)?

Original article in a Peer reviewed journal

Cheung D, Menon S, Hoare J, Dhar A, Trudgill N. Factors Associated with Upper Gastrointestinal Cancer Occurrence After Endoscopy that Did Not Diagnose Cancer. Digestive diseases and sciences 2016;61:2674-84.

Oral presentation at British Society of Gastroenterology annual meeting 2013, Glasgow

Cheung D, Menon S, Trudgill N. OC-012 How Commonly is Oesophageal Cancer Missed at Endoscopy (A UK Primary Care Based Study)? Gut 2013;62:A5-A6.

Poster presentation at Digestive Disease Week, American Gastroenterological Association annual meeting 2013, Orlando, USA

Cheung D, Evans T, Lawrence G, Trudgill N. Tu1268 How Often Is Upper Gastrointestinal Cancer Missed During Endoscopy? Gastrointestinal endoscopy 2013;77:AB481-AB.

Factors associated with upper gastro-intestinal cancer occurrence after endoscopy that did not diagnose upper gastro-intestinal cancer (an analysis of English Hospital Episode Statistics from 2003 to 2012).

Oral presentation at Digestive Disease Week, American Gastroenterological Association annual meeting 2014, Chicago, USA

Cheung D, Evison F, Patel P, Trudgill N. 452 How Commonly Is Upper Gastrointestinal Cancer Diagnosed Following an Endoscopy That Does Not Report Cancer (an Analysis of 11 Years of National Data in England)? Gastrointestinal endoscopy 2014;79:AB145-AB

Poster presentation at British Society of Gastroenterology annual meeting 2014, Manchester

Cheung D, Evison F, Patel P, Trudgill N. PTU-145 How Commonly Is Upper Gastrointestinal Cancer Diagnosed Following An Endoscopy That Does Not Report Cancer (an Analysis Of 11 Years Of National Data In England)? Gut 2014;63:A102-A

Best in stomach and duodenum section poster prize

Regional post OGD upper gastrointestinal cancer rate in the Midlands

Oral presentation at British Society of Gastroenterology annual meeting 2013, Glasgow

Cheung D, Evans T, Lawrence G, Trudgill N. OC-013 How often is upper Gastrointestinal Cancer Missed during Endoscopy? Gut 2013;62:A6-A.

Poster presentation at Digestive Disease Week, American Gastroenterological Association annual meeting 2013, Orlando, USA

Cheung D, Evans T, Lawrence G, Trudgill N. Tu1268 How Often Is Upper Gastrointestinal Cancer Missed During Endoscopy? Gastrointestinal endoscopy 2013;77:AB481-AB.

Poster presentation at British Society of Gastroenterology annual meeting 2016, Liverpool

Cheung D, Rees J, Evans T, Trudgill N. PTH-016 Factors Associated with Upper Gastrointestinal Cancer Occurrence After OGD that Did Not Detect Cancer in The West Midlands. Gut 2016;65:A225.

Oral presentation at Midland Gastroenterological Society meeting 2012

A local audit of the frequency and causes of failing to diagnose cancer at upper gastro-intestinal endoscopy.

Oral presentation at Midland Gastroenterological Society meeting 2016

Factors associated with upper gastrointestinal cancer occurrence after OGD that did not detect cancer in the West Midlands.

Best oral presentation prize

Factors associated with colorectal cancer occurrence after colonoscopy that did not diagnose colorectal cancer (an analysis of English Hospital Episode Statistics from 2003 to 2012).

Original article in a Peer reviewed journal

Cheung D, Evison F, Patel P, Trudgill N. Factors associated with colorectal cancer occurrence after colonoscopy that did not diagnose colorectal cancer. Gastrointestinal endoscopy 2016;84:287-95 e1.

Oral presentation at Digestive Disease Week, American Gastroenterological Association annual meeting 2014, Chicago, USA

Cheung D, Evison F, Patel P, Trudgill N. 168 How Commonly Is Colorectal Cancer Later Diagnosed Following a Colonoscopy That Does Not Report Colorectal Cancer (an Analysis of 11 Years of National Data in England)? Gastrointestinal endoscopy 2014;79:AB117-AB8.

Oral presentation at British Society of Gastroenterology annual meeting 2014, Manchester

Cheung D, Evison F, Patel P, Trudgill N. OC-043 How Commonly Is Colorectal Cancer Later Diagnosed Following A Colonoscopy That Does Not Report Colorectal Cancer (an Analysis Of 11 Years Of National Data In England)? Gut 2014;63:A21-A.

Oral communication prize in Colon & Anorectum section

Chapter 10: References

1 Cancer Research UK. Oesophageal cancer incidence statistics; http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence#heading-Zero [Accessed 1 October 2016].

2 Cancer Research UK. Stomach cancer incidence statistics; http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/stomach-cancer/incidence#heading-Zero [Accessed 1 October 2016].

3 Cancer Research UK. Bowel cancer incidence statistics; http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence#heading-Zero [Accessed 1 October 2016].

4 Coupland VH, Allum W, Blazeby JM, Mendall MA, Hardwick RH, Linklater KM*, et al.* Incidence and survival of oesophageal and gastric cancer in England between 1998 and 2007, a population-based study. BMC cancer 2012;**12**:11.

5 Hardikar S, Onstad L, Blount PL, Odze RD, Reid BJ, Vaughan TL. The role of tobacco, alcohol, and obesity in neoplastic progression to esophageal adenocarcinoma: a prospective study of Barrett's esophagus. PloS one 2013;**8**:e52192.

6 Lagergren J, Mattsson F, Nyren O. Gastroesophageal reflux does not alter effects of body mass index on risk of esophageal adenocarcinoma. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2014;**12**:45-51.

7 Lagergren J, Lagergren P. Oesophageal cancer. BMJ 2010;**341**:c6280.

8 Edgren G, Adami HO, Weiderpass Vainio E, Nyren O. A global assessment of the oesophageal adenocarcinoma epidemic. Gut 2013;**62**:1406-14.

9 Pohl H, Wrobel K, Bojarski C, Voderholzer W, Sonnenberg A, Rosch T*, et al.* Risk factors in the development of esophageal adenocarcinoma. The American journal of gastroenterology 2013;**108**:200-7.

10 Reim D, Loos M, Vogl F, Novotny A, Schuster T, Langer R*, et al.* Prognostic implications of the seventh edition of the international union against cancer classification for patients with gastric cancer: the Western experience of patients treated in a single-center European institution. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2013;**31**:263-71.

11 Hirschowitz BI. Endoscopic examination of the stomach and duodenal cap with the fiberscope. Lancet 1961;**1**:1074-8.

12 Amin A, Gilmour H, Graham L, Paterson-Brown S, Terrace J, Crofts TJ. Gastric adenocarcinoma missed at endoscopy. Journal of the Royal College of Surgeons of Edinburgh 2002;**47**:681-4.

13 Yalamarthi S, Witherspoon P, McCole D, Auld CD. Missed diagnoses in patients with upper gastrointestinal cancers. Endoscopy 2004;**36**:874-9.

14 Raftopoulos SC, Segarajasingam DS, Burke V, Ee HC, Yusoff IF. A cohort study of missed and new cancers after esophagogastroduodenoscopy. The American journal of gastroenterology 2010;**105**:1292-7.

15 Vradelis S, Maynard N, Warren BF, Keshav S, Travis SP. Quality control in upper gastrointestinal endoscopy: detection rates of gastric cancer in Oxford 2005-2008. Postgraduate medical journal 2011;**87**:335-9.

16 Voutilainen ME, Juhola MT. Evaluation of the diagnostic accuracy of gastroscopy to detect gastric tumours: clinicopathological features and prognosis of patients with gastric cancer missed on endoscopy. European journal of gastroenterology & hepatology 2005;**17**:1345-9.

17 Hosokawa O, Hattori M, Douden K, Hayashi H, Ohta K, Kaizaki Y. Difference in accuracy between gastroscopy and colonoscopy for detection of cancer. Hepato-gastroenterology 2007;**54**:442-4.

18 Hosokawa O, Watanabe K, Hatorri M, Douden K, Hayashi H, Kaizaki Y. Detection of gastric cancer by repeat endoscopy within a short time after negative examination. Endoscopy 2001;**33**:301-5.
19 Hosokawa O, Tsuda S, Kidani E, Watanabe K, Tanigawa Y, Shirasaki S*, et al.* Diagnosis of gastric cancer up to three years after negative upper gastrointestinal endoscopy. Endoscopy 1998;**30**:669-74.
20 Fujita S. Biology of early gastric carcinoma. Pathology, research and practice 1978;**163**:297-309.
21 International Agency for Research on Cancer WHO. *GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012;* http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx [Accessed 25 March 2015].
22 Borie F, Rigau V, Fingerhut A, Millat B, French Association for Surgical R. Prognostic factors for early gastric cancer in France: Cox regression analysis of 332 cases. World journal of surgery 2004;**28**:686-91.
23 Lassen A, Hallas J, de Muckadell OB. The risk of missed gastroesophageal cancer diagnoses in users and nonusers of antisecretory medication. Gastroenterology 2005;**129**:1179-86.
24 Chadwick G, Groene O, Hoare J, Hardwick RH, Riley S, Crosby TD*, et al.* A population-based, retrospective, cohort study of esophageal cancer missed at endoscopy. Endoscopy 2014;**46**:553-60.
25 Chadwick G, Groene O, Riley S, Hardwick R, Crosby T, Hoare J*, et al.* Gastric Cancers Missed During Endoscopy in England. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2015.
26 Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis. EIO 2014;**02**:E46-E50.
27 Lal N, Bhasin DK, Malik AK, Gupta NM, Singh K, Mehta SK. Optimal number of biopsy specimens in the diagnosis of carcinoma of the oesophagus. Gut 1992;**33**:724-6.
28 Bressler B, Paszat LF, Chen Z, Rothwell DM, Vinden C, Rabeneck L. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. Gastroenterology 2007;**132**:96-102.
29 Lee TJ, Rutter MD, Blanks RG, Moss SM, Goddard AF, Chilton A*, et al.* Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme. Gut 2012;**61**:1050-7.
30 Gavin DR, Valori RM, Anderson JT, Donnelly MT, Williams JG, Swarbrick ET. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. Gut 2013;**62**:242-9.
31 Morris EJ, Rutter MD, Finan PJ, Thomas JD, Valori R. Post-colonoscopy colorectal cancer (PCCRC) rates vary considerably depending on the method used to calculate them: a retrospective observational population-based study of PCCRC in the English National Health Service. Gut 2015;**64**:1248-56.
32 Bressler B, Paszat LF, Vinden C, Li C, He J, Rabeneck L. Colonoscopic miss rates for right-sided colon cancer: a population-based analysis. Gastroenterology 2004;**127**:452-6.
33 Erichsen R, Baron JA, Stoffel EM, Laurberg S, Sandler RS, Sorensen HT. Characteristics and survival of interval and sporadic colorectal cancer patients: a nationwide population-based cohort study. The American journal of gastroenterology 2013;**108**:1332-40.
34 Singh H, Nugent Z, Demers AA, Bernstein CN. Rate and predictors of early/missed colorectal cancers after colonoscopy in Manitoba: a population-based study. The American journal of gastroenterology 2010;**105**:2588-96.
35 Baxter NN, Sutradhar R, Forbes SS, Paszat LF, Saskin R, Rabeneck L. Analysis of administrative data finds endoscopist quality measures associated with postcolonoscopy colorectal cancer. Gastroenterology 2011;**140**:65-72.
36 Cooper GS, Xu F, Barnholtz Sloan JS, Schluchter MD, Koroukian SM. Prevalence and predictors of interval colorectal cancers in medicare beneficiaries. Cancer 2012;**118**:3044-52.

37 Morris EJ, Whitehouse LE, Farrell T, Nickerson C, Thomas JD, Quirke P*, et al.* A retrospective observational study examining the characteristics and outcomes of tumours diagnosed within and without of the English NHS Bowel Cancer Screening Programme. British journal of cancer 2012;**107**:757-64.
38 The Health Improvement Network www.epic-uk.org [Accessed 1 October 2016].
39 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases 1987;**40**:373-83.
40 Phillimore P, Beattie A, Townsend P. Widening inequality of health in northern England, 1981-91. BMJ 1994;**308**:1125-8.
41 Health & Social Care Information Centre. *Hospital Episode Statistics*; www.hscic.gov.uk/hes [Accessed 29 December 2013].
42 Shawihdi M, Thompson E, Kapoor N, Powell G, Sturgess RP, Stern N*, et al.* Variation in gastroscopy rate in English general practice and outcome for oesophagogastric cancer: retrospective analysis of Hospital Episode Statistics. Gut 2014;**63**:250-61.
43 Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S*, et al.* The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. British journal of cancer 2012;**106**:1262-7.
44 National Cancer Intelligence Network. UK Cancer e-Atlas by cancer networks.
45 Health & Social Care Information Centre. *National Bowel Cancer Audit 2009*; www.hscic.gov.uk/catalogue/PUB02587/nati-clin-audi-supp-prog-bowe-canc-2009-rep2.pdf [Accessed 2 June 2014]. 2009.
46 Health & Social Care Information Centre. *National Bowel Cancer Audit 2010*; www.hscic.gov.uk/catalogue/PUB02586/nati-clin-audi-supp-prog-bowe-canc-2010-rep.pdf [Accessed 2 June 2014]. 2010.
47 Health & Social Care Information Centre. *National Bowel Cancer Audit 2011*; www.hscic.gov.uk/catalogue/PUB02576/nati-clin-audi-supp-prog-bowe-canc-2011-rep1.pdf [Accessed 2 June 2014]. 2011.
48 Public Health England. *National Cancer Registration Service*; www.ncras.nhs.uk [Accessed 1 October 2016].
49 Health & Social Care Information Centre. *National Oesophago-Gastric Cancer Audit 2013*; www.hscic.gov.uk/catalogue/PUB11093/clin-audi-supp-prog-oeso-gast-2013-rep.pdf [Accessed 1 March 2014]. 2013.
50 Dubecz A, Gall I, Solymosi N, Schweigert M, Peters JH, Feith M*, et al.* Temporal trends in long-term survival and cure rates in esophageal cancer: a SEER database analysis. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer 2012;**7**:443-7.
51 Ruol A, Castoro C, Portale G, Cavallin F, Sileni VC, Cagol M*, et al.* Trends in management and prognosis for esophageal cancer surgery: twenty-five years of experience at a single institution. Arch Surg 2009;**144**:247-54.
52 Health & Social Care Information Centre. *National Oesophago-Gastric Cancer Audit 2012*; www.hscic.gov.uk/catalogue/PUB06331/clin-audi-supp-prog-oeso-gast-2012-rep.pdf [Accessed 1 March 2014]. 2012.
53 Rabeneck LP, L.F. Circumstances in which colonoscopy misses cancer. Frontline Gastroenterology 2010;**1**:52-8.
54 Abraham N, Barkun A, Larocque M, Fallone C, Mayrand S, Baffis V*, et al.* Predicting which patients can undergo upper endoscopy comfortably without conscious sedation. Gastrointestinal endoscopy 2002;**56**:180-9.
55 Mahajan RJ, Johnson JC, Marshall JB. Predictors of patient cooperation during gastrointestinal endoscopy. Journal of clinical gastroenterology 1997;**24**:220-3.

56 Farhadi A, Fields JZ, Hoseini SH. The assessment of esophagogastroduodenoscopy tolerance a prospective study of 300 cases. Diagnostic and therapeutic endoscopy 2001;**7**:141-7.
57 Hazeldine S, Fritschi L, Forbes G. Predicting patient tolerance of endoscopy with conscious sedation. Scandinavian journal of gastroenterology 2010;**45**:1248-54.
58 Cooper SC, Day R, Brooks C, Livings C, Thomson CS, Trudgill NJ. The influence of deprivation and ethnicity on the incidence of esophageal cancer in England. Cancer causes & control : CCC 2009;**20**:1459-67.
59 Cheung D, Evans T, Lawrence G, Trudgill N. OC-013 How often is upper Gastrointestinal Cancer Missed during Endoscopy? Gut 2013;**62**:A6-A.
60 Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. Pharmacoepidemiology and drug safety 2007;**16**:393-401.
61 Boursi B, Haynes K, Mamtani R, Yang YX. Digoxin use and the risk for colorectal cancer. Pharmacoepidemiology and drug safety 2014;**23**:1147-53.
62 Gavin AT, Francisci S, Foschi R, Donnelly DW, Lemmens V, Brenner H*, et al.* Oesophageal cancer survival in Europe: a EUROCARE-4 study. Cancer epidemiology 2012;**36**:505-12.
63 Cancer incidence for common cancers, Cancer Research UK. www.cancerresearchuk.org/cancer-info/cancerstats/incidence/commoncancers/ [Accessed 27 May 2014].
64 EUCAN, International Agency for Research on Cancer; www.eco.iarc.fr/EUCAN/ [Accessed 27 May 2014].
65 Health & Social Care Information Centre. *National Bowel Cancer Audit 2013*; www.hscic.gov.uk/catalogue/PUB11105/nati-clin-audi-supp-prog-bowe-canc-2013-rep1.pdf [Accessed 2 June 2014]. 2013.
66 Joint Advisory Group on GI Endoscopy. *BSG Quality and Safety Indicators for Endoscopy*; www.thejag.org.uk/downloads%5CUnit%20Resources%5CBSG%20Quality%20and%20Safety%20Indicators.pdf [Accessed 30 December 2013].
67 Chadwick G, Groene O, Riley S, Hardwick R, Crosby T, Hoare J*, et al.* Gastric Cancers Missed During Endoscopy in England. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2015;**13**:1264-70 e1.
68 Mulcahy HE, Kelly P, Banks MR, Connor P, Patchet SE, Farthing MJ*, et al.* Factors associated with tolerance to, and discomfort with, unsedated diagnostic gastroscopy. Scandinavian journal of gastroenterology 2001;**36**:1352-7.
69 Bodger K, Bowering K, Sarkar S, Thompson E, Pearson MG. All-cause mortality after first ERCP in England: clinically guided analysis of hospital episode statistics with linkage to registry of death. Gastrointestinal endoscopy 2011;**74**:825-33.
70 NHS digital. National Oesophago-Gastric Cancer Audits. http://content.digital.nhs.uk [Accessed 1 October 2016]
71 Macdonald S, Macleod U, Campbell NC, Weller D, Mitchell E. Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. British journal of cancer 2006;**94**:1272-80.
72 Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. British journal of cancer 2013;**108**:686-90.
73 Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. The lancet oncology 2012;**13**:353-65.
74 Kutluturkan S, Gorgulu U, Fesci H, Karavelioglu A. The effects of providing pre-gastrointestinal endoscopy written educational material on patients' anxiety: a randomised controlled trial. International journal of nursing studies 2010;**47**:1066-73.

75 Rey JF, Lambert R, Committee EQA. ESGE recommendations for quality control in gastrointestinal endoscopy: guidelines for image documentation in upper and lower GI endoscopy. Endoscopy 2001;**33**:901-3.
76 Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C*, et al.* Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. Lancet 2011;**377**:127-38.
77 Samadder NJ, Curtin K, Tuohy TM, Pappas L, Boucher K, Provenzale D*, et al.* Characteristics of missed or interval colorectal cancer and patient survival: a population-based study. Gastroenterology 2014;**146**:950-60.
78 Arain MA, Sawhney M, Sheikh S, Anway R, Thyagarajan B, Bond JH*, et al.* CIMP status of interval colon cancers: another piece to the puzzle. The American journal of gastroenterology 2010;**105**:1189-95.
79 Sawhney MS, Farrar WD, Gudiseva S, Nelson DB, Lederle FA, Rector TS*, et al.* Microsatellite instability in interval colon cancers. Gastroenterology 2006;**131**:1700-5.
80 Robertson DJ, Lieberman DA, Winawer SJ, Ahnen DJ, Baron JA, Schatzkin A*, et al.* Colorectal cancers soon after colonoscopy: a pooled multicohort analysis. Gut 2014;**63**:949-56.
81 le Clercq CM, Bouwens MW, Rondagh EJ, Bakker CM, Keulen ET, de Ridder RJ*, et al.* Postcolonoscopy colorectal cancers are preventable: a population-based study. Gut 2014;**63**:957-63.
82 Romero RV, Mahadeva S. Factors influencing quality of bowel preparation for colonoscopy. World journal of gastrointestinal endoscopy 2013;**5**:39-46.
83 Hong SN, Sung IK, Kim JH, Choe WH, Kim BK, Ko SY*, et al.* The Effect of the Bowel Preparation Status on the Risk of Missing Polyp and Adenoma during Screening Colonoscopy: A Tandem Colonoscopic Study. Clinical endoscopy 2012;**45**:404-11.
84 Shah HA, Paszat LF, Saskin R, Stukel TA, Rabeneck L. Factors associated with incomplete colonoscopy: a population-based study. Gastroenterology 2007;**132**:2297-303.
85 Tadros M, Anderson JC. Serrated polyps: clinical implications and future directions. Current gastroenterology reports 2013;**15**:342.
86 Bowles CJ, Leicester R, Romaya C, Swarbrick E, Williams CB, Epstein O. A prospective study of colonoscopy practice in the UK today: are we adequately prepared for national colorectal cancer screening tomorrow? Gut 2004;**53**:277-83.
87 Moller H, Richards S, Hanchett N, Riaz SP, Luchtenborg M, Holmberg L*, et al.* Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. British journal of cancer 2011;**105**:170-6.
88 Teh JL, Tan JR, Lau LJ, Saxena N, Salim A, Tay A*, et al.* Longer examination time improves detection of gastric cancer during diagnostic upper gastrointestinal endoscopy. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2015;**13**:480-7 e2.
89 Davies AE, Kidd D, Stone SP, MacMahon J. Pharyngeal sensation and gag reflex in healthy subjects. Lancet 1995;**345**:487-8.
90 Soma Y, Saito H, Kishibe T, Takahashi T, Tanaka H, Munakata A. Evaluation of topical pharyngeal anesthesia for upper endoscopy including factors associated with patient tolerance. Gastrointestinal endoscopy 2001;**53**:14-8.
91 Huang HH, Lee MS, Shih YL, Chu HC, Huang TY, Hsieh TY. Modified Mallampati classification as a clinical predictor of peroral esophagogastroduodenoscopy tolerance. BMC gastroenterology 2011;**11**:12.
92 Neuenschwander AU, Christensen M, Schulze S, Rosenberg J, Schulze S, Trap R. Less tachycardia during transnasal versus conventional gastroscopy. Danish medical journal 2012;**59**:A4432.

93 Yagi J, Adachi K, Arima N, Tanaka S, Ose T, Azumi T*, et al.* A prospective randomized comparative study on the safety and tolerability of transnasal esophagogastroduodenoscopy. Endoscopy 2005;**37**:1226-31.
94 Fujii T, Iishi H, Tatsuta M, Hirasawa R, Uedo N, Hifumi K*, et al.* Effectiveness of premedication with pronase for improving visibility during gastroendoscopy: a randomized controlled trial. Gastrointestinal endoscopy 1998;**47**:382-7.
95 Chang CC, Chen SH, Lin CP, Hsieh CR, Lou HY, Suk FM*, et al.* Premedication with pronase or N-acetylcysteine improves visibility during gastroendoscopy: an endoscopist-blinded, prospective, randomized study. World journal of gastroenterology : WJG 2007;**13**:444-7.
96 Bhandari P, Green S, Hamanaka H, Nakajima T, Matsuda T, Saito Y*, et al.* Use of Gascon and Pronase either as a pre-endoscopic drink or as targeted endoscopic flushes to improve visibility during gastroscopy: a prospective, randomized, controlled, blinded trial. Scandinavian journal of gastroenterology 2010;**45**:357-61.
97 Asfeldt AM, Straume B, Paulssen EJ. Impact of observer variability on the usefulness of endoscopic images for the documentation of upper gastrointestinal endoscopy. Scandinavian journal of gastroenterology 2007;**42**:1106-12.
98 Simmons DT, Harewood GC, Baron TH, Petersen BT, Wang KK, Boyd-Enders F*, et al.* Impact of endoscopist withdrawal speed on polyp yield: implications for optimal colonoscopy withdrawal time. Alimentary pharmacology & therapeutics 2006;**24**:965-71.
99 Aoun E, Abdul-Baki H, Azar C, Mourad F, Barada K, Berro Z*, et al.* A randomized single-blind trial of split-dose PEG-electrolyte solution without dietary restriction compared with whole dose PEG-electrolyte solution with dietary restriction for colonoscopy preparation. Gastrointestinal endoscopy 2005;**62**:213-8.

Appendices

Appendix 1 – West Midlands Cancer Intelligence Unit Information Request form

Information request form - please return

UKACR - United Kingdom Association of Cancer Registries

Patient Identifiable or Potentially Identifiable Data

REQUEST FORM

NB. These rules apply to England and Wales only

IMPORTANT – PLEASE READ

About this form

This form is a support document, and does not replace the requirement for a study protocol, where relevant.

It covers the release of data that are potentially identifiable or identifiable only. Do not use this form if you require non-identifiable data (see Appendix for details)

Please read and retain Guidance Notes - pages 5-6

Return form only (pages 1-4) – this form can be filled in electronically (MS Word), then printed out to sign.

Please return the completed and signed form, by post or fax only, to the address shown on the right.

Research or Audit

If requesting data for research or audit you must send a copy of your study protocol where relevant.

For research, when requesting identifiable data, you must also include a copy of the appropriate Ethics Committee approval, along with PIAG approval or evidence of patient consent.

Please send copies of all these documents with this form to the Registry

Caldicott Guardian Informed

A signed copy of this release form may be forwarded to the appropriate Caldicott Guardian

Where the recipient is based at another organisation, a further copy may be forwarded to the Caldicott Guardian of that organisation.

Declaration

Applicant should sign declaration on page 3 in all cases, but the level of identifiable data requested will determine which (if any) co-signatory is required

Data Items

Please request only the data items that you actually need for the requirements of your work. Do not request additional data items.

Although there are many data items listed, we are obliged to restrict data release to the minimum required for your study, at the highest level of anonymisation possible.

The list of data items is only intended as a starting point in a dialogue with the registry, who can help you explore the needs of your study and to inform whether particular data items are complete or robust enough for analysis

Data Interpretation

Please consult the cancer registry on the interpretation of data – for example, changes in coding practice over time can have large impacts on data, which may be artefactual and not epidemiological

Publication

If you intend to publish works containing the data in any form, please consult the cancer registry

ukacr United Kingdom Association of Cancer Registries

NHS West Midlands Cancer Intelligence Unit

SEND COMPLETED FORM TO (pages 1-4 only)

Address:	West Midlands Cancer Intelligence Unit Public Health Building The University of Birmingham Edgbaston, Birmingham B15 2TT		
Tel:		Fax:	
Email:			
Website:	www.wmpho.org.uk/wmciu		

CONTACT DETAILS

Person To Whom Data Are To Be Released	
Title	
Full Name	
Job Title	
Organisation	
Address (work)	
Postcode	
Telephone	
Fax/Mobile/Bleep	
Email	

STUDY DETAILS

Purpose

Purpose for which data are required (please explain why non-identifiable data will not meet your needs). Please attach your study protocol, if relevant.

Scope - ☐ National / ☐ Multi-Registry / ☐ Single Registry (registry involvement, please give details in text box above - all requests will be co-ordinated through the registry shown)

Information request form - please return

DATA RANGE REQUIRED

Data Requirements

☐ **Tumour site**(s) and/or ☐ **morphologies**

Sex - ☐ Male / ☐ Female / ☐ Both

Years of - ☐ Diagnosis / ☐ Death / ☐ Treatment

☐ **Geographical areas** / ☐ **Organisations** (eg Specified PCT, Electoral Ward, Trust, Hospital, Cancer Network):

Age Groups - ☐ 5 Year Groups (0-4, 5-9 ...80-84, 85+) or ☐ Other (specify) **NB. These are only for potentially identifiable data**

Any other details (please continue on page 4 if necessary)

Is there a deadline for receipt of data?

☐ YES / ☐ NO - If yes, please give date and reason:

Preferred form for receipt of data

☐ Paper / ☐ EXCEL / ☐ ASCII / ☐ Other (Specify):

DATA ITEMS REQUIRED

Discuss data requirements with the Cancer Registry

Please note that this is not meant to be an exhaustive list, but primarily a tool, intended to open a dialogue between the applicant and the Cancer Registry.

Although there may be capacity to collect all of these data items, they are not always recorded. The Cancer Registry will advise you about the completeness of these variables in relation to your request.

For example, some data items may only be available only for specific cancer sites or recent years. It may be possible to extract further details if these exist (eg. treatment types or dates). Please consult the Cancer Registry for advice.

Data Required

Please tick only the identifiers and data items that are absolutely essential for the requirements of your work

☒	*please check items required*		
Personal Details			
☐	**NHS number**		
☐	**Patient Name** (Surname, Forename, Initials)		
☐	**Surname at birth** (previous surname)		
☐	**Address** (at time of diagnosis)		
☐	**Sex**	☐	**Age** (at diagnosis)
☐	**Date of Birth**	☐	**Ethnic origin**
☐	**Postcode** (at time of diagnosis)		
	Various geographies can also be derived from postcode:		
☐	*Cancer Network*	☐	*Electoral Ward*
☐	*Strategic Health Authority*	☐	*Local Authority*
☐	*Primary Care Organisation*	☐	*ONS Super Output Area*
☐	*Government Office Region*	☐	*Deprivation Index*
Diagnostic, Tumour And Treatment Details			
☐	**Site** of primary neoplasm (or main presenting secondary when primary site is not known)		
☐	**Morphology** (type of neoplasm)		
☐	**Laterality** (side) for paired organs		
☐	**Stage** (limited information – available for breast, cervix, melanoma and colorectal cancer only)		
☐	**Grade** of tumour (degree of differentiation – available for breast and cervical cancer)		
☐	**Basis of diagnosis** (histology, cytology, haematology, clinical opinion, other tests)		
☐	**Date of diagnosis**	☐	**Year Only**
☐	**Treatment indicators** (treatment within first six months after diagnosis, where intent was curative - surgery, radiotherapy, chemotherapy, hormone, other)		
Hospital Details			
☐	**Hospital**	☐	**Trust**
☐	**Consultant** (Surname, Initials, Specialty, GMC Code)		
☐	**Unit number** (PAS Number)		
Death Details			
☐	**Alive/dead**		
☐	**Date of death**		
☐	**Cause and place of death**		
☐	**Post mortem**		

Information request form - please return

CONFIDENTIALITY / ETHICS APPROVAL

Considerations for Identifiable Data Release Only

☒ *please check all boxes that apply and give any further details in the space provided*

Patient Information Advisory Group (PIAG)

PIAG considers the use and transfer of identifiable patient data under Section 60 of the Health and Social Care Act 2001 (http://www.advisorybodies.doh.gov.uk/piag/)

Note: Identifiable data cannot be released without Ethical Committee Approval and either (A) Patient Consent or (B) PIAG approval, except to (C) authorised personnel in approved organisations– please see the Guidance Notes for further information

Type of Identifiable Data Request

Consent type obtained (please complete relevant sections)

☐ (A) Patient consent	☐ (B) PIAG approval	☐ (C) Other approved signatory

A - Individual Patient Consent

☐ I have received proof of written consent for all patients

☐ I have enclosed a copy of a blank consent form, to which all patients have agreed

B - PIAG APPROVAL (for research requests)

Has PIAG approval been given?

☐ *Yes* ☐ *No* ☐ *Application submitted*

☐ I have enclosed copies of the application and the approval letter *(essential before data can be released)*

If application submitted, please give an indication of when approval is expected:

Ethical Committee Approval (for research requests)

Has ethical committee approval been given?

☐ *Yes* ☐ *No* ☐ *Application submitted*

Type of approval sought ☐ *MREC* ☐ *LREC*

REC name

☐ I have enclosed copies of the application and the approval letter *(essential before data can be released)*

If application submitted, please give an indication of when approval is expected:

USE OF DATA

Use of Identifiable Data

Do you intend to use the data provided to contact anyone? (see note 5 of DECLARATION opposite)

☐ *Treating clinicians*	☐ *Pathologists*
☐ *Patients' GPs*	☐ *No party to be contacted*
☐ *Other (please specify) -*	

☐ I have enclosed copies of the letters to be used for contact *(essential before data can be released)*

Dissemination of results (all types of data)

Do you intend to publish your results or present the data provided in a public forum? (see notes 2 and 4 of DECLARATION opposite)

☐ *Internal meeting*	☐ *External meeting*
☐ *Internal report*	☐ *External report*
☐ *Conference*	☐ *Publication*
☐ *Other (please specify) -*	

Please give full details of the name of meeting/conference (or title of report/publication)

DECLARATION

The signature required on the declaration is dependent upon the geographical area or organisation for which the information is required. Details of this can be found in the attached APPENDIX Guidance Notes. Please contact Cancer Registry for advice.

Declaration for (potentially) Identifiable Data

I understand that, in accordance with the Data Protection Act 1998 and the UKACR Policies on the Release of Patient Identifiable Information and on the Release of Potentially Identifiable Information, data will only be released to me providing:

1. the data are only used for the purpose(s) for which they were supplied
2. the data are not passed on to other third parties unless directly concerned with their analysis or interpretation, nor will they be released into the public domain (see guidance notes - page 5)
3. any results of my/our work which are disclosed shall not be able to identify an individual
4. any results of my/our work which are disclosed into the public domain shall not show potentially identifiable information
5. no attempt will be made to identify information pertaining to particular individuals or to contact individuals, unless appropriate consent is obtained
6. the data will be kept secure at all times, will not be kept for longer than can be justified by the stated purpose, and will then be destroyed
7. no attempt will be made to link the data to other data sets, unless agreed with all data providers
8. any public domain reports or papers resulting from analyses of the provided data will be shared **prior** to publication with the Cancer Registry and appropriate acknowledgement given.

Applicant

Signature	
Date	

Designated Signatory (see appendix for guidance)

Title / Full Name	
Job Title	
Organisation	
Capacity signing under (see page 6)	
Signature	
Date	

VERIFICATION

☒ For Registry Use Only	Data Request Number
☐ Protocol received	
☐ LREC/MREC approval	PID Type of Request
☐ Patient consent seen	
☐ Caldicott Guardian Informed	
☐ PIAG approval & application number	
☐ Approved for release – signature and date	

ADDITIONAL INFORMATION

Any other information in support of your application

UKACR - United Kingdom Association of Cancer Registries

Patient Identifiable or Potentially Identifiable Data

GUIDANCE NOTES

CONFIDENTIALITY GUIDELINES

The accompanying form covers the release of data that are identifiable or potentially identifiable only. Do not use this form if you require non-identifiable data. Please return it by post or fax only.

The Need for Guidelines

To provide adequate safeguards for the individuals right to privacy at the same time as preserving the right for his/her fellows to benefit from the knowledge on cancer causation, prevention, treatment and survival that can be obtained from cancer epidemiology

Disclosure of Identifiable Data (or potentially identifiable information*) by Cancer Registries:

Regulation 2 of the Statutory Instrument (SI) on confidentiality – No. 1438, The Health Service (Control of Patient Information) Regulations 2002 – permits cancer registries to receive patient identifiable data without the need for informed consent and it permits registries to process said data for the medical purposes stipulated in regulation 2.

Data are released subject to the Confidentiality and Disclosure Guidelines, as agreed by the UKACR. Identifiable patient information (or potentially identifiable information*) will only be released to authorised personnel in legitimate organisations, as detailed in the UKACR Policy on the Release of Identifiable Patient Information (see *'Consent for release of Data'*). Full details of this policy and the guidelines covering the release of potentially identifiable information can be found under the section *'UKACR Guidelines on Confidentiality: Approved policies'* www.ukacr.org

Identifiable patient data

Include any of the following: name, address, postcode, date of birth, date or cause of death, NHS no., hospital no.

* As a general rule, the following categories should be regarded as being potentially identifiable data:

1. Individual records even if they do not include variables, such as names, full postcodes, and dates of birth which would make them obviously identifiable
2. Tabular data, based on small geographic areas**, with cell counts of fewer than five cases/events (or where counts of less than five can be inferred by simple arithmetic)
3. Tabular data containing cells that have underlying population denominators of less than 1,000

** As a general rule, the following categories should be regarded as potentially identifiable data for small geographic areas:

1. Those areas where the total denominator population is less than that of a Primary Care Trust, e.g. wards or aggregation of wards. (The smallest PCT in England has a total population of approximately 62,000 i.e. 1550 if divided into 40 single sex, 5-year age group assuming an equal size distribution)
2. Any geographic area (e.g. local authority) which, when released, may provide information regarding small population non-contiguous areas ("slivers") when combined with Primary Care Trust information. These should be regarded in the same way as ward level data
3. Any geographic area when publication in five-year age groups between 0 and 24 years is required. In this age range, particular scrutiny should be paid to tabulations and appropriate aggregations used. (Due to the rarity of cancer in children and young adults, there may be a non-negligible risk of information disclosure by for any geographic area.)

Conditions for Release of Identifiable and Potentially Identifiable Data:

Releases of both identifiable and potentially identifiable data are governed by the following principles:

1. the intended use(s) of the data should be stated clearly
2. the use(s) of the data should be justified and the data should not be used for any other purpose(s)
3. the registry should not release data that are more detailed than necessary to fulfil the stated purpose
4. the data should not be passed on to other third parties or released into the public domain***
5. the data should be kept securely for the period of time that can be justified by the stated purpose, and then destroyed
6. no attempt should be made to identify information pertaining to particular individuals or to contact individuals (unless patient consent has been obtained via the patient's clinician)
7. no attempt should be made to link the data to other data sets, unless agreed with the data providers
8. any public domain*** reports or papers resulting from analyses of the provided data should be shared prior to publication with the cancer registry (or registries) supplying the information.
9. recipients of data should be aware of their responsibilities, and should sign an agreement to this effect prior to the release of data by a registry

***Public Domain – Definition

Publication of data on a website and in unrestricted circulations of reports or documents containing data should be regarded as being in the public domain.

UKACR POLICY

UKACR Policy - www.ukacr.org

*The following section has been extracted from **Appendix 2** of the UKACR Policy on the Release of Identifiable Patient Information follows (for full details see UKACR Guidelines on Confidentiality at* www.ukacr.org*)*

Policies for Data Release – Full Documents

Full details covering the policies on the release of identifiable information and the release of potentially identifiable information are available from the UKACR website under section *UKACR Guidelines on Confidentiality: Approved Policies* in www.ukacr.org). These policies are part of the documentation submitted to and subject to annual review by the Patient Information Advisory Group (PIAG: **http://www.advisorybodies.doh.gov.uk/piag/**).

PIAG was established to provide advice on issues of national significance involving the use of patient information and to oversee arrangements created under Section 60 of the Health and Social Care Act 2001.

CONSENT FOR RELEASE OF DATA

Appropriate Signatories - Designated Individuals

Wherever possible, a **registered health professional** should sign requests for the cancer registration information. We recognise that people with other training are appointed to some of these posts; for example, Directors of Public Health for some Primary Care Trusts, where they could reasonably be expected to sign *ex officio*.

Possible signatories for the different organisations requesting different types of data are listed below, but their inclusion does not necessarily mean that anyone or only one of them can sign for a given request. Depending on the purpose of the request, **Cancer Registry Directors** should use their discretion as to who should sign, and for some requests, several signatories may be needed. For some purposes, signatories may be specified in Service Level Agreements between registries and particular organisations.

CONSENT FOR RELEASE OF DATA

Appropriate Signatories - Designated Individuals

Table to help you discern the most appropriate people to sign your form, depending on your organisation type and nature of data requested.

Type of Data	Possible Signatories
NHS Trust	
Clinician's own data or data regarding patients of predecessor	Clinician (needs to sign stating taken over care of patients from predecessor) Caldicott Guardian (for patients not now managed by any clinician in trust) Medical Director (for patients not now managed by any clinician in trust)
Cancer Site-Specific data	Lead Cancer Clinician Lead Tumour Site-Specific Clinician Trust's Lead Clinician for Audit (if data request stated for audit) Caldicott Guardian Medical Director
Data for the whole Trust	Lead Cancer Clinician Trust's Lead Clinician for Audit (if data request stated for audit) Caldicott Guardian Medical Director
Data for Split Site Trusts / Cancer Centres	Lead Cancer Clinician from each unit Trust's Lead Clinician for Audit from each unit (if stated for audit) Caldicott Guardian from each unit Medical Director from each unit
Private Hospital	
Clinician's own data	Clinician
All hospital / unit: only the data they sent us if possible (assumed purpose is audit)	Signatures of all Clinicians of patients involved Medical Director
All NHS Trusts they serve for pathology cases only, and only the data they sent us if possible	Senior Pathologist/ Clinical Head of Pathology Service Caldicott Guardian from each of the Trusts Medical Directors from each of the Trusts
Private Pathology Laboratory	
Own pathology patients only, and only the data they sent us if possible	Head of Pathology Service
Hospice	
Own patients – complete records	Signatures of all Clinicians Medical Director
GP Practice	
Own patients – either at diagnosis or registered at time of request	Lead GP for Cancer Signatures of all GPs Head of Practice Caldicott Guardian of relevant PCT
Quality Assurance Reference Centres	
Relevant patients – screening age groups	Covered by Service Level Agreement
Cancer Network	
All Network – either patients diagnosed or treated in the network or for cases resident in the network geographical boundaries	Network Lead Cancer Clinician Network's Lead Clinician for Audit (if data request stated for audit)
Multi-Disciplinary Team – either patients diagnosed or treated in the network or for cases resident in the network geographical boundaries	Network Lead Cancer Clinician Network Lead Tumour Site-Specific Clinician Network's Lead Clinician for Audit (if data request stated for audit) Signatures of all Clinicians in the MDT Lead Clinician for the MDT
Cancer site-specific data – either patients diagnosed or treated in the Network or for cases resident in Network geographical boundaries	Network Lead Cancer Clinician Network Lead tumour Site-Specific Clinician Network's Lead Clinician for Audit (if data request stated for audit)
Primary Care Trusts	
Own Patients – by GP practice register or geographical boundaries	Director of Public Health (DPH) Caldicott Guardian
Consortium analysis arrangements where one PCT is doing some work on behalf of others (say in an area previously covered by a single HA)	Signatures of all DsPH of the participating PCTs Caldicott Guardian of all DsPH of the participating PCTs

CONSENT FOR DEATH INFORMATION

Release of Date and Cause of Death Information

Where possible a registered health professional should sign requests for death information, as this information forms part of the complete cancer registration record.

Type of Data	Possible Signatories
Audit data	Signatory should be the Lead Clinician for Audit for the unit.
Clinical Trials data	Signatory should be the leading investigator within the unit (the signatory is actually signing that the patient had given informed consent to take part in the trial and to be followed-up, as part of the research trial).
Medical Records data	Signatory should be the Lead Cancer Clinician for the unit.
Cancer Site-Specific Audit or any other Audits	When clinicians are requesting information on their own patients, or as part of a Trust or Network site-specific audit, and are providing information on the patients concerned such as names & dates of birth, then release of information is allowed without obtaining the individual signatures.

Appendix 2 - Gastric cancer Read codes

Readcode	Description
B11..00	Malignant neoplasm of stomach
B11..11	Gastric neoplasm
B110.00	Malignant neoplasm of cardia of stomach
B110000	Malignant neoplasm of cardiac orifice of stomach
B110111	Malignant neoplasm of gastro-oesophageal junction
B110z00	Malignant neoplasm of cardia of stomach NOS
B111.00	Malignant neoplasm of pylorus of stomach
B111000	Malignant neoplasm of prepylorus of stomach
B111100	Malignant neoplasm of pyloric canal of stomach
B111z00	Malignant neoplasm of pylorus of stomach NOS
B112.00	Malignant neoplasm of pyloric antrum of stomach
B113.00	Malignant neoplasm of fundus of stomach
B114.00	Malignant neoplasm of body of stomach
B115.00	Malignant neoplasm of lesser curve of stomach unspecified
B116.00	Malignant neoplasm of greater curve of stomach unspecified
B117.00	Malignant neoplasm, overlapping lesion of stomach
B119.00	Siewert type III adenocarcinoma
B11y.00	Malignant neoplasm of other specified site of stomach
B11y000	Malignant neoplasm of anterior wall of stomach NEC
B11y100	Malignant neoplasm of posterior wall of stomach NEC
B11yz00	Malignant neoplasm of other specified site of stomach NOS
B11z.00	Malignant neoplasm of stomach NOS

Appendix 3 – Oesophageal cancer Read codes

Readcode	Description
B10..00	Malignant neoplasm of oesophagus
B100.00	Malignant neoplasm of cervical oesophagus
B101.00	Malignant neoplasm of thoracic oesophagus
B102.00	Malignant neoplasm of abdominal oesophagus
B103.00	Malignant neoplasm of upper third of oesophagus
B104.00	Malignant neoplasm of middle third of oesophagus
B105.00	Malignant neoplasm of lower third of oesophagus
B107.00	Siewert type I adenocarcinoma
B10y.00	Malignant neoplasm of other specified part of oesophagus
B10z.00	Malignant neoplasm of oesophagus NOS
B10z.11	Oesophageal cancer
B905000	Neoplasm of uncertain behaviour of oesophagus
B110100	Malignant neoplasm of cardio-oesophageal junction of stomach
B110111	Malignant neoplasm of gastro-oesophageal junction
B106.00	Malignant neoplasm, overlapping lesion of oesophagus
B118.00	Siewert type II adenocarcinoma

Appendix 4 – ICD-10 codes for upper gastrointestinal cancer sites

Proximal oesophagus

ICD-10 code	Description
C15.0	Cervical part of oesophagus
C15.1	Thoracic part of oesophagus
C15.3	Upper third of oesophagus
C15.4	Middle third of oesophagus

Distal oesophagus

C15.2	Abdominal part of oesophagus
C15.5	Lower third of oesophagus

Oesophagus unspecified

C15.8	Overlapping lesion of oesophagus
C15.9	Oesophagus, unspecified

Proximal stomach

C16.0	Cardia and gastro-oesophageal junction
C16.1	Fundus of stomach

Mid stomach

C16.2	Body of stomach
C16.5	Lesser curvature of stomach, unspecified
C16.6	Greater curvature of stomach, unspecified

Distal stomach

C16.3	Gastric antrum

C16.4 Pylorus

Stomach unspecified

C16.8 Overlapping lesion of stomach

C16.9 Stomach, unspecified

Appendix 5 – ICD-10 codes for colorectal cancer sites

ICD-10 code for right sided CRC	Description
C18.0	Caecum, Ileocaecal valve
C18.2	Ascending colon
C18.3	Hepatic flexure
C18.4	Transverse colon

ICD-10 code for left sided CRC	Description
C18.5	Splenic flexure
C18.6	Descending colon
C18.7	Sigmoid colon
C19	Rectosigmoid junction
C20	Rectum

ICD-10 code for unspecified CRC location	Description
C18.8	Overlapping lesion of colon
C18.9	Colon, unspecified

Appendix 6 - ICD-10 codes for colorectal polyps

ICD-10 code	Description
D12.0	Caecal polyp(s)
D12.2	Ascending colon polyp(s)
D12.3	Transverse colon, hepatic flexure, splenic flexure polyp(s)
D12.4	Descending colon polyp(s)
D12.5	Sigmoid colon polyp(s)
D12.6	Colon, site unspecified polyp(s)
D12.7	Rectosigmoid junction polyp(s)
D12.8	Rectal polyp(s)

Appendix 7 - ICD-10 codes for metastases

ICD-10 code	Description
C77.1	Intrathoracic lymph nodes
C77.2	Intra-abdominal lymph nodes
C77.4	Inguinal and lower limb lymph nodes
C77.5	Intrapelvic lymph nodes
C78.0	Secondary malignant neoplasm of lung
C78.6	Secondary malignant neoplasm of retroperitoneum and peritoneum
C78.7	Secondary malignant neoplasm of liver and intrahepatic bile duct
C79.5	Secondary malignant neoplasm of bone and bone marrow
C34	Malignant neoplasm of bronchus and lung
C48	Malignant neoplasm of retroperitoneum and peritoneum
C22	Malignant neoplasm of liver
C40-C41	Malignant neoplasms of bone and articular cartilage

Appendix 8 - OPCS-4 codes for UGIC surgical procedures

OPCS-4 code	Description
G011	Oesophagogastrectomy and anastomosis of oesophagus to stomach, Ivor Lewis procedure
G012	Oesophagogastrectomy and anastomosis of oesophagus to transposed jejunum
G013	Oesophagogastrectomy and anastomosis of oesophagus to jejunum NEC
G018	Excision of oesophagus and stomach, Other specified
G019	Excision of oesophagus and stomach, Unspecified
G021	Total oesophagectomy and anastomosis of pharynx to stomach
G022	Total oesophagectomy and interposition of microvascularly attached jejunum
G023	Total oesophagectomy and interposition of jejunum NEC
G024	Total oesophagectomy and interposition of microvascularly attached colon
G025	Total oesophagectomy and interposition of colon NEC
G028	Total excision of oesophagus, other specified
G029	Total excision of oesophagus, Unspecified
G031	Partial oesophagectomy and end to end anastomosis of oesophagus
G032	Partial oesophagectomy and interposition of microvascularly attached jejunum
G033	Partial oesophagectomy and anastomosis of oesophagus to transposed jejunum
G034	Partial oesophagectomy and anastomosis of oesophagus to jejunum NEC
G035	Partial oesophagectomy and interposition of microvascularly attached colon
G036	Partial oesophagectomy and interposition of colon NEC

G038	Partial excision of oesophagus, other specified
G039	Oesophagectomy Unspecified, Partial excision of oesophagus, Unspecified
G041	Excision of lesion of oesophagus
G042	Open laser destruction of lesion of oesophagus
G043	Open destruction of lesion of oesophagus NEC
G048	Open extirpation of lesion of oesophagus, Other specified
G049	Open extirpation of lesion of oesophagus, Unspecified
G271	Total gastrectomy and excision of surrounding tissue
G272	Total gastrectomy and anastomosis of oesophagus to duodenum
G273	Total gastrectomy and interposition of jejunum
G274	Total gastrectomy and anastomosis of oesophagus to transposed jejunum
G275	Total gastrectomy and anastomosis of oesophagus to jejunum NEC
G278	Total excision of stomach, other specified
G279	Total gastrectomy, Unspecified
G281	Partial gastrectomy and anastomosis of stomach to duodenum
G282	Partial gastrectomy and anastomosis of stomach to transposed jejunum
G283	Partial gastrectomy and anastomosis of stomach to jejunum NEC, Billroth II
G284	Sleeve gastrectomy and duodenal switch
G285	Sleeve gastrectomy NEC
G288	Partial excision of stomach, other specified
G289	Gastrectomy NEC
G291	Open excision of polyp of stomach
G292	Open excision of lesion of stomach NEC
G293	Open laser destruction of lesion of stomach

G298	Open extirpation of lesion of stomach, other specified
G299	Unspecified open extirpation of lesion of stomach

Appendix 9 - OPCS-4 codes for chemotherapy

OPCS-4 code	Description
X70	Procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X71	Procurement of drugs for chemotherapy for neoplasm in Bands 6-10
X72	Delivery of Chemotherapy for neoplasm
X73	Delivery of oral chemotherapy for neoplasm
X352	Intravenous chemotherapy
X384	Subcutaneous chemotherapy
X373	Intramuscular chemotherapy
Z082	Follow up examination after chemotherapy for malignant neoplasm
Z511	Chemotherapy session for neoplasm
Z542	Convalescence following chemotherapy

Appendix 10 - OPCS-4 codes for radiotherapy

OPCS-4 code	Description
X652	Delivery of a fraction of intracavitary radiotherapy
X653	Delivery of a fraction of interstitial radiotherapy
X654	Delivery of a fraction of external beam radiotherapy NEC
X658	Other specified radiotherapy delivery
X659	Unspecified radiotherapy delivery
Y902	Radiotherapy NEC
Y914	Megavoltage treatment for adaptive radiotherapy
Y911	Delivery of a fraction of complex radiotherapy on a megavoltage machine
Y912	Delivery of a fraction of simple radiotherapy on a megavoltage machine
Y913	Delivery of a fraction of radiotherapy on a superficial or orthovoltage machine
Y918	Other specified radiotherapy procedures
Y919	Unspecified radiotherapy procedures

Appendix 11 - OPCS-4 codes for CRC surgical procedures

OPCS-4 code	Description
H04	Total excision of colon and rectum
H05	Total excision of colon
H06	Extended excision of right hemicolon
H07	Other excision of right hemicolon
H08	Excision of transverse colon
H09	Excision of left hemicolon
H10	Excision of sigmoid colon
H11	Other excision of colon
H29	Subtotal excision of colon
H33	Excision of rectum
H40	Operations on rectum through anal sphincter
H122	Excision of lesion of colon NEC
H123	Destruction of lesion of colon NEC
H128	Other specified extirpation of lesion of colon
H129	Unspecified extirpation of lesion of colon
H341	Open excision of lesion of rectum
H345	Open destruction of lesion of rectum
H348	Other specified open extirpation of lesion of rectum
H349	Unspecified open extirpation of lesion of rectum
H402	Trans-sphincteric excision of lesion of rectum
H403	Trans-sphincteric destruction of lesion of rectum

Appendix A – SQL codes for POUGIC (chapter 5)

```
---creating a temporary table with all endoscopies performed regardless of whether people have had a
diagnosis of upper GI cancer and information associated with them

select a.Patient_ID, a.SPELL_ID, a.SPELL_FINYR, a.EPIORDER, a.ADMIDATE, a.DISDATE, a.ADMIMETH,
CLASSPAT, a.DIAG_01, DIAG_02, DIAG_03, DIAG_04,DIAG_05, DIAG_06,
DIAG_07, DIAG_08,DIAG_09, DIAG_10,DIAG_11, DIAG_12, DIAG_13, DIAG_14, DISMETH, GPPRAC, MAINSPEF,
a.OPERTN_01, OPERTN_02, OPERTN_03, OPERTN_04, OPERTN_05,
OPERTN_06, OPERTN_07, opertn_08, OPERTN_09, OPERTN_10, OPERTN_11, OPERTN_12, OPERTN_13, OPERTN_14,
LEFT(procode,3) procode, SEX, STARTAGE, quintiles , resgor, b.op_date as endoscopy_date
into #endoscopies
from HED_BASETables.dbo.HEs_EPISODES a inner join HED_BASETABLES.dbo.HES_OPER_Extract b
                                                    on a.Epikey = b.epikey
                                                    and a.patient_ID = b.patient_ID
                                                    and a.spell_ID = b.spell_ID
where left(oper_code,3) in  ('G16','G45')  ---these are the endoscopy codes used
                                  and ADMIDATE < '20120401' and admidate> '20030331' and a.pt_spellyear
in (2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013)

-----(4711739 row(s) affected)

---now identifiying patients with a diagnosis of upper GI cancer, based on inpatient admissions
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  a.Patient_ID, MIN(a.epistart) minadmission, ICD10
into #UpperGIDiagnosis
from HED_Basetables.dbo.HES_EPISODES a inner join HED_BAsetables.dbo.HES_Diag_extract b
                                  on a.Epikey = b.epikey
                                                    and a.patient_ID = b.patient_ID
                                                    and a.spell_ID = b.spell_ID
where  left(ICD10,3) in ('C15','C16') and a.pt_spellyear in
(2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015)
group by a.PAtient_ID, ICD10

----319309
---Looking at deaths with a code of upper GI Cancer
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  Extract_HESID, Max(DOD) dod
into #UpperGIDiagnosisfromDeaths
from (select Extract_HESID, DOD
        from HES_DW_ONS.dbo.HED_ONS_Mortality
        where
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON_NEONATA
L_11+CAUSE_OF_DEATH_NON_NEONATAL_12+CAUSE_OF_DEATH_NON_NEONATAL_13
        +CAUSE_OF_DEATH_NON_NEONATAL_14+CAUSE_OF_DEATH_NON_NEONATAL_15+CAUSE_OF_DEATH_NON_NEONATAL_2+C
AUSE_OF_DEATH_NON_NEONATAL_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+
        CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUS
E_OF_DEATH_NON_NEONATAL_9 like '%c1[56]%'
                ) t1
group by Extract_HESID
-----(171254 row(s) affected)
---checking to see whether there are any patients who died of Upper GI Cancer who weren't admitted to
hosp with it
select *
from  #UpperGIDiagnosisfromDeaths t1  left join #UpperGIDiagnosis t2
                                                    on t1.EXTRACT_HESID = t2.Patient_ID collate
database_default
where t2.Patient_ID is null
---there are 19103 patients here who died without a hospital admission owhere upper gi cancer was
recorded
---Have a look to see whether there is anyone who visited outpatients with an upper GI diagnosis
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  PatientID,diag_01, MIN(Apptdate) minadmission
into #UpperGIDiagnosis_OP
from (select PatientID, Apptdate, diag_01
        from HES_DW.dbo.HES_OP_200607
        where  diag_01 like 'c1[56]%'
        union All
        select PatientID , Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_200708
        where diag_01 like 'c1[56]%'
                union All
```

```
        select PatientID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_200809
        where diag_01 like 'c1[56]%'
                union ALl
        select PatientID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_200910
        where diag_01 like 'c1[56]%'
                union ALl
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201011
        where diag_01 like 'c1[56]%'
                union ALl
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201112
        where diag_01 like 'c1[56]%'
                        union ALl
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201213
        where diag_01 like 'c1[56]%'

        union ALl
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201314
        where diag_01 like 'c1[56]%'
        union ALl
        select Patient_ID collate database_default, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201415
        where diag_01 like 'c1[56]%'
        ) t1
group by PAtientID, diag_01
----(14631 row(s) affected)

---Look to see whether there are any OP diagnoses which are not in IP
select *
from  #UpperGIDiagnosis_OP t1  left join #UpperGIDiagnosis t2
                                                        on t1.PAtientID = t2.Patient_ID collate
database_default
where t2.Patient_ID is null
---3395
---collating diagnoses from all sources and identifying the first records with upperGI cancer

select Patient_ID, MIN(minadmission)minadmission
Into #UpperGI
from (select Patient_ID, minadmission
                from #UpperGIDiagnosis
                union all
                select EXTRACT_HESID collate SQL_Latin1_General_CP1_CI_AS, DOD
                from #UpperGIDiagnosisfromDeaths
                union all
                select PatientID collate SQL_Latin1_General_CP1_CI_AS, minadmission
                from #UpperGIDiagnosis_OP) a
group by Patient_ID
----234097

--truncate table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses;
---drop table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses;

---creating a table which only has the endoscopy info for the patients who have had a diagnosisi of
upper gi cancer at some point
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct t1.*, t2.minadmission firstAdmissionDiagnosis
into [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
from #endoscopies t1 inner join #UpperGI t2
                                on t1.Patient_ID= t2.PAtient_ID collate database_default

---(250565 row(s) affected)
----looking at the data
select *
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
---having looked at this it has confirmed that there are patients with endoscopys after their cancer
admission dates, we need to remove these rows
----checking whether there are any endoscopies performed after the diagnosis of upper gi cancer as we
won't be interested in these
select *
```

```
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where ADMIDATE > firstAdmissionDiagnosis
---91923 rows
---and now deleting them
delete
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where ADMIDATE > firstAdmissionDiagnosis
----(91923 row(s) affected)

---having a look at records where the endoscopy does not occur during the endoscopy admission
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where endoscopy_date not between ADMIDATE and  disdate
---(310 row(s) affected)
---updating these cases so the endoscopy occurs on discharge date
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set endoscopy_date  = disdate
where endoscopy_date not between ADMIDATE and  disdate
---(310 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set endoscopy_date  = disdate
where endoscopy_date  is null
---151

---adding in some additional variables that we will need for these analyses
alter table  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add dayspostopEndscopyDiag int,
LastEndoscopybeforeDiag int,
MissedDiag int;

---calculating the length of time from the endoscopy until the diagnosis of upper GI cancer - as this
will be needed to know whether the
---endoscopy was successful or not

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set dayspostopEndscopyDiag = DATEDIFF(dd, endoscopy_date, firstAdmissionDiagnosis);
---(158642 row(s) affected)
---Lets check out the results of this
select dayspostopEndscopyDiag, COUNT(*)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
group by dayspostopEndscopyDiag
order by dayspostopEndscopyDiag

---there are some nulls and also some negative numbers
select SPELL_FINYR, ADMIDATE, endoscopy_date, firstAdmissionDiagnosis, dayspostopEndscopyDiag
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where dayspostopEndscopyDiag is null or dayspostopEndscopyDiag < 0
order by Spell_finYr
--the nulls look to be where the endoscopy date is null
---these look all to be for the admitting epsiode as the diagnosis has been put onto the admission
fate,lets look to see wherther
----any of these people will be
select SPELL_FINYR, ADMIDATE, DISDATE, endoscopy_date, firstAdmissionDiagnosis, dayspostopEndscopyDiag
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where ADMIDATE >  firstAdmissionDiagnosis and( dayspostopEndscopyDiag is null or
dayspostopEndscopyDiag < 0)
order by Spell_finYr


---Should may update the diagnosis date to the discharge date? but first looking in some more detail
at the dates
select SPELL_FINYR, ADMIDATE, disdate, endoscopy_date, firstAdmissionDiagnosis, dayspostopEndscopyDiag
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where dayspostopEndscopyDiag is null or dayspostopEndscopyDiag < 0
order by Spell_finYr

---lets leave this as it is now and replace the nulls - using the discharge date as the endoscopy date

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set dayspostopEndscopyDiag = DATEDIFF(dd, disdate, firstAdmissionDiagnosis)
where endoscopy_date is null
---28 rows
---there are some disdates here which are null and should be updated
```

```
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set DISDATE = t2.disdate
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_SPELLS t2
                                                        on t1.Patient_ID = t2.PATIENT_ID
                                                        and t1.SPELL_ID = t2.SPELL_ID
where t1.DISDATE is null
-----(22651 row(s) affected)

---lets for back to the endoscopy date and look to see whether any of the negative dates occur within
3 months of the diagnosis  - this woiuld have been considered a success
select  SPELL_FINYR, ADMIDATE, disdate, endoscopy_date, firstAdmissionDiagnosis,  case when
ABS(dayspostopEndscopyDiag) >182.5 then 1 else 0 end,dayspostopEndscopyDiag
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where dayspostopEndscopyDiag is null or dayspostopEndscopyDiag < 0
order by case when ABS(dayspostopEndscopyDiag) > 182.5 then 1 else 0 end
---22272 rows

 ---there are 8 rows where the endoscopy date and the admission date are more than 182.5 days apart
but in the same spell
 --lets have a look at these patients see what information is available o them

select Patient_ID, SPELL_ID, EPIORDER,  ADMIDATE, DISDATE, endoscopy_date, firstAdmissionDiagnosis,
dayspostopEndscopyDiag
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where dayspostopEndscopyDiag < -182.5
Order by Patient_ID
---this has returned 2 rows - I think that some of these will have to be removed, as the endoscopy has
occurred months after the original admission date, but we won't know
---when the diagnosis occurred

--maybe creat a drop record field now
alter table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add droprecord int;
---and drop the 2 records that have been found

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where dayspostopEndscopyDiag  < - 182.5
----(0 row(s) affected) - we cannot be sure when during their admission they were diagnosed with
oesopahgeal cancer
---Lets look at the patient IDs for these people
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where Patient_ID in ('4FAA6521022B0D5C13B18EDE67872633','DCA2A874E0986333C5804C42A114AC6D')

---adding a nother variable to the data - diagnosis this variable will be a one where the endoscopy
occurred within 90 days of the diagnosis
alter table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add diagnosis int;

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set diagnosis =1
where dayspostopEndscopyDiag between 0 and  180
----(129494 row(s) affected)
---and set the diagnosis = 1 where the diagnosis ocurred in the same spell as the endoscopy
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and diagnosis is null and dayspostopEndscopyDiag < 0
--21632
---and update these records
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set diagnosis =1
where droprecord is null and diagnosis is null and dayspostopEndscopyDiag < 0
-----(21632 row(s) affected)

---lets update the missed diagnoses
Update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set MissedDiag = 1
where dayspostopEndscopyDiag between 181 and 1826
 ----(32266 row(s) affected)

 ---check this is doing what it should be
 select *
```

```
 from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
 where MissedDiag = 1

 ----I need to update the discharge method to that of the spells

update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set DISMETH = t2.dismeth
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_SPElls t2
                                                                    on t1.Patient_ID =
t2.PATIENT_ID
                                                                    and t1.SPELL_ID =
t2.SPELL_ID
where pt_spellyear in (2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2013,2014,2015)

-----(158470 row(s) affected)

select MissedDiag, count(*)
from   [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
group by MissedDiag
---lets look at cleaning the data - first checking the genders are correct

select sex, COUNT(patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by sex

---there are 38 people with an invalid sex
---lets look at these patietns
select Patient_ID, admidate
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where SEX in (0,9)
---38 rows

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord =1
where SEX in (0,9)
----(11 row(s) affected)
---moving on to checking the GOR codes are withing England
select resgor, COUNT(distinct patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by resgor
order by resgor

---there are 52 patients here of no fixed  abode
---there are in total 700 patients with a GOR which is not valid, lets look at these patients and see
whether they have been admitted
--- another time in the for the same groups with a fixed address
select resgor, COUNT(distinct Patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where Patient_ID in (select Patient_ID
                                        from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                        where resgor in ('u','x','y','z')
                                        )
         and resgor in ('a','b','c','d','e','f','g','h','i','k','j')
group by resgor
order by resgor
---about 100 are...
select distinct Patient_ID
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where Patient_ID in (select Patient_ID
                                        from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                        where resgor in ('u','x','y','z')
                                        )
         and resgor in ('a','b','c','d','e','f','g','h','i','k','j')
---104 rows
----    lets have a look at these patients
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where Patient_ID in (select distinct Patient_ID
```

```
                                                from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                where Patient_ID in (select Patient_ID
                                                                                                                from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                                                                                where resgor
in ('u','y','z')
                                                                                                                )
                                                                                        and resgor in
('a','b','c','d','e','f','g','h','i','k','j'))
order by Patient_ID, admidate
---312
---lets leave the 'u's as they are and update the y's and z's



----I think we'll have to drop the records where a patient has been listed as living in Scotland,
Wales or abroad as they may have had other admissions in these countries
---that are not recorded in HES
select distinct Patient_ID
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where resgor in ('s','w','x')
---797 rows

update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in (select distinct Patient_ID
                                        from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                        where resgor in ('s','w','x'))
----(807 row(s) affected)

select distinct t1.PAtient_ID, t1.resgor, t2.resgor
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
Patient_ID, resgor

                                        from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses

                                        where resgor in
('a','b','c','d','e','f','g','h','i','k','j')) t2
                                                                                                                on
t1.patient_ID = t2.patient_ID
where t1.resgor in ('y','z')
order by PAtient_ID
---89 rows - check these are disinct patients

select distinct patient_ID
from (select distinct t1.PAtient_ID, t1.resgor, t2.resgor resgornew
                        from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
(select distinct Patient_ID, resgor

                                                                        from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses

                                                                        where resgor in
('a','b','c','d','e','f','g','h','i','k','j')) t2

                on t1.patient_ID = t2.patient_ID
                        where t1.resgor in ('y','z'))  a
---89 so this is correct
update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set resgor  = t2.resgornew
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
t1.PAtient_ID, t1.resgor, t2.resgor resgornew

                                                from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct Patient_ID,
resgor



                                                                                                                from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
                                                                    where resgor
in ('a','b','c','d','e','f','g','h','i','k','j')) t2

                                        on t1.patient_ID = t2.patient_ID

                                                where t1.resgor in ('y','z')) t2
                                                                    on t1.Patient_ID=
t2.PAtient_ID
                                                                    and t1.resgor =
t2.resgor
-----(75 row(s) affected)

---lets relook at the data now
select resgor, COUNT(distinct patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord  is null
group by resgor
order by resgor
---there are still  269 patients with an invalid GOR code


---and update
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set resgor = t2.resgornew
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
t1.Patient_ID, t1.resgor resgornew, t2.resgor

                                                from HED_Basetables.dbo.HES_Episodes t1 inner join
(select resgor, Patient_ID, Admidate

                                    from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses

                                    where droprecord  is null and resgor in ('y','z')) t2

                                                                                        on
t1.PATIENT_ID = t2.Patient_ID

        and abs(DATEDIFF(dd, t1.Admidate, t2.Admidate)) between 0 and 90

                                                where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k') and PT_SpellYear in
(2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015)) t2
                                                                    on t1.Patient_ID =
t2.Patient_ID
                                                                    and t1.resgor =
t2.resgor
----(97row(s) affected)
--and now check how this has worked
select resgor, COUNT(distinct patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord  is null
group by resgor
order by resgor
---down to 179 - lets look at outpatients for these patietns

---(18 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set resgor = t2.resgornew
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
t1.Patient_ID, t1.resgor resgornew, t2.resgor

                                                from HED_basetables.dbo.HES_Episodes t1 inner join
(select resgor, Patient_ID, Admidate

                                    from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                                    on
t1.PATIENT_ID = t2.Patient_ID



                                                    where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k') and PT_SpellYear in
(2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2016)) t2
                                                                                    on t1.Patient_ID =
t2.Patient_ID
                                                                                    and t1.resgor =
t2.resgor
----(55 row(s) affected)
---these ones can't be updated so may need to be removed

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where resgor in ('y','z')
----(54 row(s) affected)
---check the spell year
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where SPELL_FINYR = '9999'
---(10 row(s) affected)

-----Look at admimeth
select admimeth, COUNT(distinct patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord  is null
group by admimeth
order by admimeth

----not sure what to do with these records -  I don't suppose it will matter too much
---check disdate
select disdate, COUNT(distinct patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord  is null
group by disdate
order by disdate
---this looks OK
---move onto dismeth
select dismeth, COUNT(distinct patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord  is null
group by dismeth
order by dismeth
---there are 31 here with validation errors - I'm not sure what to do with these either!
---lets look at the class pats
select classpat, COUNT(distinct patient_ID)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord  is null
group by classpat
order by classpat
---we've got some regular day attenders here ....
---lets loook at these patients
select *
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and CLASSPAT =3
order by Patient_ID

---127 rows

----Lets have a look at the number of people admitted by the year of their first admission
select COUNT(distinct Patient_ID) , case when firstAdmissionDiagnosis between convert(datetime,
'2001/04/01') and CONVERT(datetime, '2002/03/31') then '2001/02'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/04/01') and CONVERT(datetime, '2003/03/31')
then '2002/03'
```

```
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/04/01') and CONVERT(datetime, '2004/03/31')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/04/01') and CONVERT(datetime, '2005/03/31')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/04/01') and CONVERT(datetime, '2006/03/31')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/04/01') and CONVERT(datetime, '2007/03/31')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/04/01') and CONVERT(datetime, '2008/03/31')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/04/01') and CONVERT(datetime, '2009/03/31')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/04/01') and CONVERT(datetime, '2010/03/31')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/04/01') and CONVERT(datetime, '2011/03/31')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/04/01') and CONVERT(datetime, '2012/03/31')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/04/01') and CONVERT(datetime, '2013/03/31')
then '2012/13' end fin_yr
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by case when firstAdmissionDiagnosis between convert(datetime, '2001/04/01') and
CONVERT(datetime, '2002/03/31') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/04/01') and CONVERT(datetime, '2003/03/31')
then '2002/03'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/04/01') and CONVERT(datetime, '2004/03/31')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/04/01') and CONVERT(datetime, '2005/03/31')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/04/01') and CONVERT(datetime, '2006/03/31')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/04/01') and CONVERT(datetime, '2007/03/31')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/04/01') and CONVERT(datetime, '2008/03/31')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/04/01') and CONVERT(datetime, '2009/03/31')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/04/01') and CONVERT(datetime, '2010/03/31')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/04/01') and CONVERT(datetime, '2011/03/31')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/04/01') and CONVERT(datetime, '2012/03/31')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/04/01') and CONVERT(datetime, '2013/03/31')
then '2012/13' end
order by case when firstAdmissionDiagnosis between convert(datetime, '2001/04/01') and
CONVERT(datetime, '2002/03/31') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/04/01') and CONVERT(datetime, '2003/03/31')
then '2002/03'
```

```
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/04/01') and CONVERT(datetime, '2004/03/31')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/04/01') and CONVERT(datetime, '2005/03/31')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/04/01') and CONVERT(datetime, '2006/03/31')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/04/01') and CONVERT(datetime, '2007/03/31')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/04/01') and CONVERT(datetime, '2008/03/31')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/04/01') and CONVERT(datetime, '2009/03/31')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/04/01') and CONVERT(datetime, '2010/03/31')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/04/01') and CONVERT(datetime, '2011/03/31')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/04/01') and CONVERT(datetime, '2012/03/31')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/04/01') and CONVERT(datetime, '2013/03/31')
then '2012/13' end
---we've identified 24118 patients disgnosed in 2006/07 and 2007/08 combined
---there are also a few nulls here - so we probably need to check this out as wll
---Lets have a look at the nulls from earlier
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and  case when firstAdmissionDiagnosis between convert(datetime,
'2001/04/01') and CONVERT(datetime, '2002/03/31') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/04/01') and CONVERT(datetime, '2003/03/31')
then '2002/03'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/04/01') and CONVERT(datetime, '2004/03/31')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/04/01') and CONVERT(datetime, '2005/03/31')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/04/01') and CONVERT(datetime, '2006/03/31')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/04/01') and CONVERT(datetime, '2007/03/31')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/04/01') and CONVERT(datetime, '2008/03/31')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/04/01') and CONVERT(datetime, '2009/03/31')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/04/01') and CONVERT(datetime, '2010/03/31')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/04/01') and CONVERT(datetime, '2011/03/31')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/04/01') and CONVERT(datetime, '2012/03/31')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/04/01') and CONVERT(datetime, '2013/03/31')
then '2012/13' end  is null
order by Patient_ID, admidate
---these are all spells for which  the admission occurred in March 2001, but discharge occurred after
then
---lets have a look to see whether we are dropping records in one year more than in any other
```

```
Select case when firstAdmissionDiagnosis between convert(datetime, '2001/04/01') and CONVERT(datetime,
'2002/03/31') then '2001/02'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/04/01') and CONVERT(datetime, '2003/03/31')
then '2002/03'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/04/01') and CONVERT(datetime, '2004/03/31')
then '2003/04'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/04/01') and CONVERT(datetime, '2005/03/31')
then '2004/05'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/04/01') and CONVERT(datetime, '2006/03/31')
then '2005/06'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/04/01') and CONVERT(datetime, '2007/03/31')
then '2006/07'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/04/01') and CONVERT(datetime, '2008/03/31')
then '2007/08'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/04/01') and CONVERT(datetime, '2009/03/31')
then '2008/09'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/04/01') and CONVERT(datetime, '2010/03/31')
then '2009/10'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/04/01') and CONVERT(datetime, '2011/03/31')
then '2010/11'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/04/01') and CONVERT(datetime, '2012/03/31')
then '2011/12'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/04/01') and CONVERT(datetime, '2013/03/31')
then '2012/13' end fin_yr,
                                                                                    sum(droprecord)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
group by case when firstAdmissionDiagnosis between convert(datetime, '2001/04/01') and
CONVERT(datetime, '2002/03/31') then '2001/02'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/04/01') and CONVERT(datetime, '2003/03/31')
then '2002/03'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/04/01') and CONVERT(datetime, '2004/03/31')
then '2003/04'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/04/01') and CONVERT(datetime, '2005/03/31')
then '2004/05'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/04/01') and CONVERT(datetime, '2006/03/31')
then '2005/06'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/04/01') and CONVERT(datetime, '2007/03/31')
then '2006/07'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/04/01') and CONVERT(datetime, '2008/03/31')
then '2007/08'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/04/01') and CONVERT(datetime, '2009/03/31')
then '2008/09'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/04/01') and CONVERT(datetime, '2010/03/31')
then '2009/10'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/04/01') and CONVERT(datetime, '2011/03/31')
then '2010/11'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/04/01') and CONVERT(datetime, '2012/03/31')
then '2011/12'
                                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/04/01') and CONVERT(datetime, '2013/03/31')
then '2012/13' end
---nope, this seems to be consistent
```

```
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null

----Lets add dod and causes of death to this table
alter table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add  timetodeath int,
underlying varchar(5),
cause_1 varchar(5),
cause_2 varchar(5),
cause_3 varchar(5);

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set timetodeath = datediff(month,firstAdmissionDiagnosis , dod)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Extract_HeSID,
MAX(DOD) DOD

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality

                                        group by EXTRACT_HESID) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
----(130875 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set underlying = t2.CAUSE_OF_DEATH
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,

                                                t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD


                                    from HES_DW_ONS.dbo.HED_ONS_Mortality


                                    group by EXTRACT_HESID) t2

                                        on t1.EXTRACT_HESID = t2.EXTRACT_HESID

                                        and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
----(130875 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set cause_1 = t2.CAUSE_OF_DEATH_NON_NEONATAL_1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,

                                                t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD


                                    from HES_DW_ONS.dbo.HED_ONS_Mortality


                                    group by EXTRACT_HESID) t2

                                        on t1.EXTRACT_HESID = t2.EXTRACT_HESID

                                        and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
---(130875 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set cause_2 = t2.CAUSE_OF_DEATH_NON_NEONATAL_2
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,
```

```
                                        t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                    from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD

                                from HES_DW_ONS.dbo.HED_ONS_Mortality

                                group by EXTRACT_HESID) t2

                                    on t1.EXTRACT_HESID = t2.EXTRACT_HESID

                                    and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
----    (130875 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set cause_3 = t2.CAUSE_OF_DEATH_NON_NEONATAL_3
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,

                                        t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                    from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD

                                from HES_DW_ONS.dbo.HED_ONS_Mortality

                                group by EXTRACT_HESID) t2

                                    on t1.EXTRACT_HESID = t2.EXTRACT_HESID

                                    and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
-----   (130875 row(s) affected)

----lets have a look at the data I've just added in
select Patient_ID, SPELL_FINYR, ADMIDATE, DISDATE,dismeth,firstAdmissionDiagnosis,
timetodeath,underlying, cause_1, cause_2, cause_3
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where timetodeath is not null


---check whether there are any deaths recorded as happening before the first record of GI cancer
select Patient_ID, SPELL_FINYR, ADMIDATE, DISDATE,dismeth,firstAdmissionDiagnosis,
timetodeath,underlying, cause_1, cause_2, cause_3
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where timetodeath is not null    and timetodeath < 0
--- 18 rows, and most of these the DOD is a  long time before the cancer aignosis
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where  timetodeath is not null   and timetodeath < 0
-----18 row(s) affected)

--lets just check the ages of these paitnets
select startAge, COUNT(distinct Patient_ID)

               from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by StartAgE
order by StartAgE

---27 patietns don't seem to have a startage, there are also 20 patients under 16 (including 5 who are
under 1)
---Lets look at the patients who do not have a start age
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
where STARTAGE is null
---119 rows
---quite a lot of these have a drop record
select *
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where STARTAGE is null  and droprecord is null
---only 39 rows here
select Patient_ID, StartAge
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where Patient_ID in (select distinct Patient_ID
                                                from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                where STARTAGE is null  and droprecord is null )
---no additional age groups here...
---Lets have a look to see whether there are any ages for these patietns on HES

----lets update these now...
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set STARTAGE = Admiage
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select t1.Patient_ID,
t1.StartAge, ADMIAGE, ACTIVAGE, ENDAGE

                                        from HED_Basetables.dbo.HES_EPISODES t1 inner join (select
distinct Patient_ID, spell_ID


                                    from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses


                                    where STARTAGE is null  and droprecord is null )t2

                                                                                        on t1.Patient_ID =
t2.Patient_ID

                                                                                        and t1.SPELL_ID =
t2.SPELL_ID) t2
                                                                                        on t1.Patient_ID =
t2.Patient_ID
where t1.STARTAGE is null
----(32 row(s) affected)

---have a look in the first hes episodes tab;e


update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where STARTAGE is null or STARTAGE = 999
----(27 row(s) affected)
---after speaking to Ravi he suggested that the best hting to do would be to remove those patients
under 18
update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in (select distinct Patient_ID
                                        from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                        where STARTAGE < 18 or STARTAGE > 200)
-----(47 row(s) affected)



---Need to add some additional factors here looking at chemo/radiotherapy/surgery and ethnicity
alter table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add chemo int,
radiotherapy int,
Surgery varchar(50),
ethnos varchar(1),
Renal_disease int,
Acute_MI int,
Cerebral_vasc_acc int,
CHF int,
Conn_tissue_disorder int,
Dementia int,
Diabetes int,
Liver_disease int,
```

```
Peptic_Ulcer int,
Peripheral_vasc_disease int,
Pulmonary_disease int,
Cancer int,
Diabetes_complications int,
Paraplegia int,
Metastatic_cancer int,
HIV int,
Severe_liver_dis int;
---lets update the records to indicate wehtehr or not people have had chemo - this would probably be
better done inan individual table
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 0 ;


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
OP_DATE

                                        from HED_Basetables.dbo.HES_OPER_extract

                                        where OPer_code like 'x7[0123]%'

                                                or OPer_code in ( 'X352', 'X384', 'X373'))t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.OP_DATE
----(44686 row(s) affected)
----- and the first part of the HES tabl

----lets  look to see if we can get any more infor off of the ICD 10 codes
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
EPISTART

                                        from HED_Basetables.dbo.HES_diag_extract

                                        where icd10 in ('z082' , 'z511', 'z542')) t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.EPISTART
---46943
---there are some here where there isn't a record of chemo



update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Surgery  = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
OP_DATE

                                        from HED_Basetables.dbo.HES_OPER_extract

                                        where OPER_CODE like 'G0[123]%' or OPER_CODE like 'G2[78]%'

                                        ) t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.OP_DATE
---(35684 row(s) affected)



----updating the comorbid conditions
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Acute_MI = case when diag_01 like 'I2[123]%' or diag_02 like 'I2[123]%' or diag_03 like 'I2[123]%'
or diag_04 like 'I2[123]%'  or diag_05 like 'I2[123]%'
```

```
                or diag_06 like 'I2[123]%'  or diag_07 like 'I2[123]%'  or diag_08 like 'I2[123]%'
or diag_09 like 'I2[123]%'  or diag_10 like 'I2[123]%'
                 or diag_11 like 'I2[123]%'  or diag_12 like 'I2[123]%'  or diag_13 like 'I2[123]%'
or diag_14 like 'I2[123]%'   then 1
        WHEN  diag_01 like 'I25[28]%' or diag_02 like 'I25[28]%' or diag_03 like 'I25[28]%' or diag_04
like 'I25[28]%' or diag_05 like 'I25[28]%' or diag_06 like 'I25[28]%' or
        diag_07 like 'I25[28]%' or diag_08 like 'I25[28]%' or diag_09 like 'I25[28]%' or diag_10 like
'I25[28]%' or diag_11 like 'I25[28]%' or diag_12 like 'I25[28]%' or
        diag_13 like 'I25[28]%' or diag_14 like 'I25[28]%'        then 1 else 0 end ;
-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cerebral_vasc_acc = case when diag_01 like 'I6%' or diag_02 like 'I6%' or diag_03 like 'I6%'  or
diag_04 like 'I6%'  or diag_05 like 'I6%'
                or diag_06 like 'I6%'  or diag_07 like 'I6%'  or diag_08 like 'I6%'  or diag_09 like
'I6%'  or diag_10 like 'I6%'
                 or diag_11 like 'I6%'  or diag_12 like 'I6%'  or diag_13 like 'I6%'  or diag_14 like
'I6%'  then 1
        WHEN  diag_01 like 'G45[012489]%' or diag_02 like 'G45[012489]%' or diag_03 like
'G45[012489]%' or diag_04 like 'G45[012489]%' or diag_05 like 'G45[012489]%' or diag_06 like
'G45[012489]%' or
                diag_07 like 'G45[012489]%' or diag_08 like 'G45[012489]%' or diag_09 like
'G45[012489]%' or diag_10 like 'G45[012489]%' or diag_11 like 'G45[012489]%' or diag_12 like
'G45[012489]%' or
                diag_13 like 'G45[012489]%' or diag_14 like 'G45[012489]%' then 1
        WHEN  diag_01 like 'G46%' or diag_02 like 'G46%' or diag_03 like 'G46%' or diag_04 like 'G46%'
or diag_05 like 'G46%' or diag_06 like 'G46%' or
                diag_07 like 'G46%' or diag_08 like 'G46%' or diag_09 like 'G46%' or diag_10 like
'G46' or diag_11 like 'G46%' or diag_12 like 'G46%' or
                diag_13 like 'G46%' or diag_14 like 'G46%'        then 1 else 0 end;
-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set CHF = case when diag_01 like 'I50%' or diag_02 like 'I50%' or diag_03 like 'I50%'  or diag_04 like
'I50%'  or diag_05 like 'I50%'
                 or diag_06 like 'I50%'  or diag_07 like 'I50%'  or diag_08 like 'I50%'  or diag_09
like 'I50%'  or diag_10 like 'I50%'
                  or diag_11 like 'I50%'  or diag_12 like 'I50%'  or diag_13 like 'I50%'  or diag_14
like 'I50%'  then 1
         else 0 end ;
        -----(195126 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Conn_tissue_disorder = case when diag_01 like 'M05%' or diag_02 like 'M05%' or diag_03 like 'M05%'
or diag_04 like 'M05%'  or diag_05 like 'M05%'
                 or diag_06 like 'M05%'  or diag_07 like 'M05%'  or diag_08 like 'M05%'  or diag_09
like 'M05%'  or diag_10 like 'M05%'
                  or diag_11 like 'M05%'  or diag_12 like 'M05%'  or diag_13 like 'M05%'  or diag_14
like 'M05%'   then 1
        WHEN  diag_01 like 'M06[039]%' or diag_02 like 'M06[039]%' or diag_03 like 'M06[039]%' or
diag_04 like 'M06[039]%' or diag_05 like 'M06[039]%' or diag_06 like 'M06[039]%' or
                diag_07 like 'M06[039]%' or diag_08 like 'M06[039]%' or diag_09 like 'M06[039]%' or
diag_10 like 'M06[039]%' or diag_11 like 'M06[039]%' or diag_12 like 'M06[039]%' or
                diag_13 like 'M06[039]%' or diag_14 like 'M06[039]%'  then 1
        WHEN  diag_01 like 'M3[24]' or diag_02 like 'M3[24]%' or diag_03 like 'M3[24]%' or diag_04
like 'M3[24]%' or diag_05 like 'M3[24]%' or diag_06 like 'M3[24]%' or
                diag_07 like 'M3[24]%' or diag_08 like 'M3[24]%' or diag_09 like 'M3[24]%' or diag_10
like 'M3[24]' or diag_11 like 'M3[24]%' or diag_12 like 'M3[24]%' or
                diag_13 like 'M3[24]%' or diag_14 like 'M3[24]%'  then 1
        When    diag_01 = 'M332' or diag_02 = 'M332' or diag_03= 'M332' or diag_04 = 'M332' or diag_05
= 'M332' or diag_06 = 'M332' or
                diag_07 = 'M332'or diag_08 = 'M332' or diag_09 = 'M332' or diag_10 = 'M332' or diag_11
= 'M332' or diag_12 = 'M332' or
                diag_13 = 'M332'or diag_14 = 'M332'       then 1
        When    diag_01 = 'M353' or diag_02 = 'M353' or diag_03= 'M353' or diag_04 = 'M353' or diag_05
= 'M353' or diag_06 = 'M353' or
                diag_07 = 'M353'or diag_08 = 'M353' or diag_09 = 'M353' or diag_10 = 'M353' or diag_11
= 'M353' or diag_12 = 'M353' or
                diag_13 = 'M353'or diag_14 = 'M353'  then 1              else 0 end ;
-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Dementia = case when diag_01 like 'F0[0123]%' or diag_02 like 'F0[0123]%' or diag_03 like
'F0[0123]%'  or diag_04 like 'F0[0123]%'  or diag_05 like 'F0[0123]%'
                 or diag_06 like 'F0[0123]%'  or diag_07 like 'F0[0123]%'  or diag_08 like 'F0[0123]%'
or diag_09 like 'F0[0123]%'  or diag_10 like 'F0[0123]%'
```

```
                or diag_11 like 'F0[0123]%'  or diag_12 like 'F0[0123]%'  or diag_13 like
'F0[0123]%'  or diag_14 like 'F0[0123]%'  then 1
        WHEN  diag_01 like 'F051%' or diag_02 like 'F051%' or diag_03 like 'F051%' or diag_04 like
'F051%' or diag_05 like 'F051%' or diag_06 like 'F051%' or
                diag_07 like 'F051%' or diag_08 like 'F051%' or diag_09 like 'F051%' or diag_10 like
'F051%' or diag_11 like 'F051%' or diag_12 like 'F051%' or
                diag_13 like 'F051%' or diag_14 like 'F051%'      then 1 else 0 end ;
-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diabetes = case when diag_01 like 'E1[0134][15689]%' or diag_02 like 'E1[0134][15689]%' or diag_03
like 'E1[0134][15689]%'  or diag_04 like 'E1[0134][15689]%'  or diag_05 like 'E1[0134][15689]%'
                or diag_06 like 'E1[0134][15689]%'  or diag_07 like 'E1[0134][15689]%'  or diag_08
like 'E1[0134][15689]%'  or diag_09 like 'E1[0134][15689]%'  or diag_10 like 'E1[0134][15689]%'
                or diag_11 like 'E1[0134][15689]%'  or diag_12 like 'E1[0134][15689]%'  or diag_13
like 'E1[0134][15689]%'  or diag_14 like 'E1[0134][15689]%'  then 1 else 0 end;
-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Liver_disease =  case when diag_01 like 'K70[23]%' or diag_02 like 'K70[23]%' or diag_03 like
'K70[23]%'  or diag_04 like 'K70[23]%'  or diag_05 like 'K70[23]%'
                or diag_06 like 'K70[23]%'  or diag_07 like 'K70[23]%'  or diag_08 like 'K70[23]%'
or diag_09 like 'K70[23]%'  or diag_10 like 'K70[23]%'
                or diag_11 like 'K70[23]%'  or diag_12 like 'K70[23]%'  or diag_13 like 'K70[23]%'
or diag_14 like 'K70[23]%'    then 1
        WHEN  diag_01 like 'K717%' or diag_02 like 'K717%' or diag_03 like 'K717%' or diag_04 like
'K717%' or diag_05 like 'K717%' or diag_06 like 'K717%' or
                diag_07 like 'K717%' or diag_08 like 'K717%' or diag_09 like 'K717%' or diag_10 like
'K717%' or diag_11 like 'K717%' or diag_12 like 'K717%' or
                diag_13 like 'K717%' or diag_14 like 'K717%'      then 1
        when diag_01 like 'K7[34]%' or diag_02 like 'K7[34]%' or diag_03 like 'K7[34]%'  or diag_04
like 'K7[34]%'  or diag_05 like 'K7[34]%'
                or diag_06 like 'K7[34]%'  or diag_07 like 'K7[34]%'  or diag_08 like 'K7[34]%'  or
diag_09 like 'K7[34]%'  or diag_10 like 'K7[34]%'
                or diag_11 like 'K7[34]%'  or diag_12 like 'K7[34]%'  or diag_13 like 'K7[34]%'  or
diag_14 like 'K7[34]%' then 1 else 0 end ;

-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Peptic_Ulcer = case when diag_01 like 'k2[5678]%' or diag_02 like 'k2[5678]%' or diag_03 like
'k2[5678]%'  or diag_04 like 'k2[5678]%'  or diag_05 like 'k2[5678]%'
                or diag_06 like 'k2[5678]%'  or diag_07 like 'k2[5678]%'  or diag_08 like 'k2[5678]%'
or diag_09 like 'k2[5678]%'  or diag_10 like 'k2[5678]%'
                or diag_11 like 'k2[5678]%'  or diag_12 like 'k2[5678]%'  or diag_13 like
'k2[5678]%'  or diag_14 like 'k2[5678]%'  then 1 else 0 end;

-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Peripheral_vasc_disease = case when diag_01 in ('I739','I790','z958', 'z959') or diag_02 in
('I739','I790','z958', 'z959')or diag_03 in ('I739','I790','z958', 'z959')  or diag_04 in
('I739','I790','z958', 'z959') or diag_05 in ('I739','I790','z958', 'z959')
                or diag_06 in ('I739','I790','z958', 'z959')  or diag_07 in ('I739','I790','z958',
'z959')  or diag_08 in ('I739','I790','z958', 'z959')  or diag_09 in ('I739','I790','z958', 'z959')
or diag_10 in ('I739','I790','z958', 'z959')
                or diag_11 in ('I739','I790','z958', 'z959')  or diag_12 in ('I739','I790','z958',
'z959')  or diag_13 in ('I739','I790','z958', 'z959') or diag_14 in ('I739','I790','z958', 'z959')
then 1
        WHEN  diag_01 like 'I71%' or diag_02 like 'I71%' or diag_03 like 'I71%' or diag_04 like 'I71%'
or diag_05 like 'I71%' or diag_06 like 'I71%' or
                diag_07 like 'I71%' or diag_08 like 'I71%' or diag_09 like 'I71%' or diag_10 like
'I71%' or diag_11 like 'I71%' or diag_12 like 'I71%' or
                diag_13 like 'I71%' or diag_14 like 'I71%'        then 1
        when diag_01 like 'r02%' or diag_02 like 'r02%' or diag_03 like 'r02%'  or diag_04 like 'r02%'
or diag_05 like 'r02%'
                or diag_06 like 'r02%'  or diag_07 like 'r02%'  or diag_08 like 'r02%'  or diag_09
like 'r02%'  or diag_10 like 'r02%'
                or diag_11 like 'r02%'  or diag_12 like 'r02%'  or diag_13 like 'r02%'  or diag_14
like 'r02%' then 1 else 0 end ;

-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Pulmonary_disease = case     WHEN  diag_01 like 'J4[01234567]%' or diag_02 like 'J4[01234567]%' or
diag_03 like 'J4[01234567]%' or diag_04 like 'J4[01234567]%' or diag_05 like 'J4[01234567]%' or
diag_06 like 'J4[01234567]%' or
```

```
                diag_07 like 'J4[01234567]%' or diag_08 like 'J4[01234567]%' or diag_09 like
'J4[01234567]%' or diag_10 like 'J4[01234567]%' or diag_11 like 'J4[01234567]%' or diag_12 like
'J4[01234567]%' or
                diag_13 like 'J4[01234567]%' or diag_14 like 'J4[01234567]%'  then 1
        when diag_01 like 'J6[01234567]%' or diag_02 like 'J6[01234567]%' or diag_03 like
'J6[01234567]%'  or diag_04 like 'J6[01234567]%'  or diag_05 like 'J6[01234567]%'
                 or diag_06 like 'J6[01234567]%'  or diag_07 like 'J6[01234567]%'  or diag_08 like
'J6[01234567]%'  or diag_09 like 'J6[01234567]%'  or diag_10 like 'J6[01234567]%'
                  or diag_11 like 'J6[01234567]%'  or diag_12 like 'J6[01234567]%'  or diag_13 like
'J6[01234567]%'  or diag_14 like 'J6[01234567]%'   then 1 else 0 end ;

-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cancer = case       WHEN  diag_01 like 'c[012345689]%' or diag_02 like 'C[01234567]%' or diag_03
like 'c[012345689]%' or diag_04 like 'c[012345689]%' or diag_05 like 'c[012345689]%' or diag_06 like
'c[012345689]%' or
                diag_07 like 'c[012345689]%' or diag_08 like 'c[012345689]%' or diag_09 like
'c[012345689]%' or diag_10 like 'c[012345689]%' or diag_11 like 'c[012345689]%' or diag_12 like
'c[012345689]%' or
                diag_13 like 'c[012345689]%' or diag_14 like 'c[012345689]%'       then 1
        when diag_01 like 'c7[0123456]%' or diag_02 like 'c7[0123456]%' or diag_03 like 'c7[0123456]%'
or diag_04 like 'c7[0123456]%'  or diag_05 like 'c7[0123456]%'
                 or diag_06 like 'c7[0123456]%'  or diag_07 like 'c7[0123456]%'  or diag_08 like
'c7[0123456]%'  or diag_09 like 'c7[0123456]%'  or diag_10 like 'c7[0123456]%'
                  or diag_11 like 'c7[0123456]%'  or diag_12 like 'c7[0123456]%'  or diag_13 like
'c7[0123456]%'  or diag_14 like 'c7[0123456]%'   then 1 else 0 end;

-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diabetes_complications =  case when diag_01 like 'E1[0134][2347]%' or diag_02 like
'E1[0134][2347]%' or diag_03 like 'E1[0134][2347]%'  or diag_04 like 'E1[0134][2347]%'  or diag_05
like 'E1[0134][2347]%'
                 or diag_06 like 'E1[0134][2347]%'  or diag_07 like 'E1[0134][2347]%'  or diag_08 like
'E1[0134][2347]%'  or diag_09 like 'E1[0134][2347]%'  or diag_10 like 'E1[0134][2347]%'
                  or diag_11 like 'E1[0134][2347]%'  or diag_12 like 'E1[0134][2347]%'  or diag_13
like 'E1[0134][2347]%'  or diag_14 like 'E1[0134][2347]%'    then 1 else 0 end ;

-----(195126 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Paraplegia =case when diag_01 in ('G041',  'G820', 'G821', 'G822') or diag_02 in ('G041',  'G820',
'G821', 'G822')or diag_03 in ('G041',  'G820', 'G821', 'G822')  or diag_04 in ('G041',  'G820',
'G821', 'G822') or diag_05 in ('G041',  'G820', 'G821', 'G822')
                 or diag_06 in ('G041',  'G820', 'G821', 'G822')  or diag_07 in ('G041',  'G820',
'G821', 'G822')  or diag_08 in ('G041',  'G820', 'G821', 'G822')  or diag_09 in ('G041',  'G820',
'G821', 'G822')  or diag_10 in ('G041',  'G820', 'G821', 'G822')
                  or diag_11 in ('G041',  'G820', 'G821', 'G822')  or diag_12 in ('G041',  'G820',
'G821', 'G822')  or diag_13 in ('G041',  'G820', 'G821', 'G822') or diag_14 in ('G041',  'G820',
'G821', 'G822')  then 1
        WHEN  diag_01 like 'G81%' or diag_02 like 'G81%' or diag_03 like 'G81%' or diag_04 like 'G81%'
or diag_05 like 'G81%' or diag_06 like 'G81%' or
                diag_07 like 'G81%' or diag_08 like 'G81%' or diag_09 like 'G81%' or diag_10 like
'G81%' or diag_11 like 'G81%' or diag_12 like 'G81%' or
                diag_13 like 'G81%' or diag_14 like 'G81%' then 1         else 0 end ;

-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Metastatic_cancer =case      WHEN  diag_01 like 'c7[789]%' or diag_02 like 'c7[789]%' or diag_03
like 'c7[789]%' or diag_04 like 'c7[789]%' or diag_05 like 'c7[789]%' or diag_06 like 'c7[789]%' or
                diag_07 like 'c7[789]%' or diag_08 like 'c7[789]%' or diag_09 like 'c7[789]%' or
diag_10 like 'c7[789]%' or diag_11 like 'c7[789]%' or diag_12 like 'c7[789]%' or
                diag_13 like 'c7[789]%' or diag_14 like 'c7[789]%'  then 1         else 0 end ;

-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set HIV=  case  WHEN  diag_01 like 'B2[01234]%' or diag_02 like 'B2[01234]%' or diag_03 like
'B2[01234]%' or diag_04 like 'B2[01234]%' or diag_05 like 'B2[01234]%' or diag_06 like 'B2[01234]%' or
                diag_07 like 'B2[01234]%' or diag_08 like 'B2[01234]%' or diag_09 like 'B2[01234]%' or
diag_10 like 'B2[01234]%' or diag_11 like 'B2[01234]%' or diag_12 like 'B2[01234]%' or
                diag_13 like 'B2[01234]%' or diag_14 like 'B2[01234]%'     then 1  else 0 end ;
-----(195126 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
set Severe_liver_dis=  case when diag_01 in ('k721',  'k729', 'k766', 'k767') or diag_02 in ('k721',
'k729', 'k766', 'k767')or diag_03 in ('k721',  'k729', 'k766', 'k767') or diag_04 in ('k721',  'k729',
'k766', 'k767') or diag_05 in ('k721',  'k729', 'k766', 'k767')
                or diag_06 in ('k721',  'k729', 'k766', 'k767')  or diag_07 in ('k721',  'k729',
'k766', 'k767')  or diag_08 in ('k721',  'k729', 'k766', 'k767')  or diag_09 in ('k721',  'k729',
'k766', 'k767')  or diag_10 in ('k721',  'k729', 'k766', 'k767')
                 or diag_11 in ('k721',  'k729', 'k766', 'k767')  or diag_12 in ('k721',  'k729',
'k766', 'k767')  or diag_13 in ('k721',  'k729', 'k766', 'k767') or diag_14 in ('k721',  'k729',
'k766', 'k767')  then 1 else 0 end ;
-----(195126 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Renal_disease =  case       WHEN  diag_01 like 'I1[23]%' or diag_02 like 'I1[23]%' or diag_03
like 'I1[23]%' or diag_04 like 'I1[23]%' or diag_05 like 'I1[23]%' or diag_06 like 'I1[23]%' or
               diag_07 like 'I1[23]%' or diag_08 like 'I1[23]%' or diag_09 like 'I1[23]%' or diag_10
like 'I1[23]%' or diag_11 like 'I1[23]%' or diag_12 like 'I1[23]%' or
               diag_13 like 'I1[23]%' or diag_14 like 'I1[23]%'  then 1
       when diag_01 like 'N0[13]%' or diag_02 like 'N0[13]%' or diag_03 like 'N0[13]%'  or diag_04
like 'N0[13]%'  or diag_05 like 'N0[13]%'
                or diag_06 like 'N0[13]%'  or diag_07 like 'N0[13]%'  or diag_08 like 'N0[13]%'  or
diag_09 like 'N0[13]%'  or diag_10 like 'N0[13]%'
                 or diag_11 like 'N0[13]%'  or diag_12 like 'N0[13]%'  or diag_13 like 'N0[13]%'  or
diag_14 like 'N0[13]%'    then 1
       when diag_01 like 'N1[89]%' or diag_02 like 'N1[89]%' or diag_03 like 'N1[89]%'  or diag_04
like 'N1[89]%'  or diag_05 like 'N1[89]%'
                or diag_06 like 'N1[89]%'  or diag_07 like 'N1[89]%'  or diag_08 like 'N1[89]%'  or
diag_09 like 'N1[89]%'  or diag_10 like 'N1[89]%'
                 or diag_11 like 'N1[89]%'  or diag_12 like 'N1[89]%'  or diag_13 like 'N1[89]%'  or
diag_14 like 'N1[89]%'    then 1
        when diag_01 like 'N25%' or diag_02 like 'N25%' or diag_03 like 'N25%'  or diag_04 like
'N25%'  or diag_05 like 'N25%'
                or diag_06 like 'N25%'  or diag_07 like 'N25%'  or diag_08 like 'N25%'  or diag_09
like 'N25%'  or diag_10 like 'N25%'
                 or diag_11 like 'N25%'  or diag_12 like 'N25%'  or diag_13 like 'N25%'  or diag_14
like 'N25%'  then 1
       when diag_01 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074') or diag_02 in
('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')or diag_03 in ('N052', 'N053', 'n054',
'N055', 'N056', 'N072', 'N073', 'N074')  or diag_04 in ('N052', 'N053', 'n054', 'N055', 'N056',
'N072', 'N073', 'N074') or diag_05 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')
                or diag_06 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or
diag_07 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or diag_08 in ('N052',
'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or diag_09 in ('N052', 'N053', 'n054',
'N055', 'N056', 'N072', 'N073', 'N074')  or diag_10 in ('N052', 'N053', 'n054', 'N055', 'N056',
'N072', 'N073', 'N074')
                 or diag_11 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or
diag_12 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or diag_13 in ('N052',
'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074') or diag_14 in ('N052', 'N053', 'n054', 'N055',
'N056', 'N072', 'N073', 'N074')  then 1
                  else 0 end;
-----(195126 row(s) affected)



---check how many cases of og cnacer we identified to match that seen in th epaper the paper
identified 23521 patients

---there are some missing endoscopy dates still
update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set endoscopy_date = DISDATE
where ADMIDATE = DISDATE and endoscopy_date is null
---0
update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set endoscopy_date = ADMIDATE
where  endoscopy_date is null
---(55 row(s) affected)


---upodate those records where the endoscopy date is outside the period we are looking at
update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
where  endoscopy_date  <  '20030104'
----(45763 row(s) affected)
update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set droprecord = 1
```

```
where  endoscopy_date  >  '20120331'
----53



---Need to add ethnicity to these tables
update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set ethnos = t2.ethnos
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_BAsetables.dbo.HES_EPISODES t2
                                                        on t1.Patient_ID = t2.Patient_ID
                                                        and t1.SPELL_ID = t2.SPELL_ID
                                                        and t1.EPIORDER = t2.EPIORDER
where PT_SpellYear in (2002, 2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013)

-----(158642 row(s) affected)



---lets look at these ethnicities
select ethnos, COUNT(*)
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by ethnos
order by ethnos
---there are quite  a  lot here which are invalid
---lets have a look whether there are any better ethnicities and any patietns who only have one
ethnicity recorded
select A.Patient_ID, A.Ethnos, A.admidate
into #t3
from HED_basetables.dbo.HES_EPISODES a inner join
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses b
                                                ON A.patient_id = b.Patient_ID
where  PT_SpellYear in (2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014) and
A.ETHNOS not in ('0','1','2','3','4','5','6','7','8','9','x','y','z')


----2200761
select  t1.patient_ID
into #t2
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1  inner join #t3 t2

        on t1.Patient_ID = t2.Patient_ID
where t1.ethnos  in ('0','1','2','3','4','5','6','7','8','9','x','y','z')
group by t1.Patient_ID
having COUNT(distinct t2.ethnos) =1
----12226 rows
---updating these patietns
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set ethnos = t3.ETHNOS
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join #t2   t2

                on t1.Patient_ID = t2.PAtient_ID

                                        inner join #t3 t3
                                                                                                on
t1.Patient_ID = t3.PAtient_ID
---(20932 row(s) affected)
select patient_ID, ethnos, count(*) admissions
into #t4
from #t3
group by patient_ID, ethnos

Update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set ethnos = t2.ETHNOS
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join #t4   t2

                on t1.Patient_ID = t2.PAtient_ID

                                        inner join (select PAtient_ID, max(admissions)admissions

                                                        from #t4

                                                        group by Patient_ID) t3
```

```
                on t2.Patient_ID = t3.Patient_ID

                and t2.admissions = t3.admissions

where  t1.Patient_ID not in  (select Patient_ID from #t2)  and  t1.ethnos  in
('0','1','2','3','4','5','6','7','8','9','x','y','z')
---1166

---lets check out the ethnicites again
select ethnos, COUNT(*)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
group by ethnos
order by ethnos
---this is a huge improvement on the previous data - still some old codes though...

alter table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add  ethnic_group varchar(50);

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set ethnic_group = case     when ethnos in ('A','B','C','0') then 'White'
                                        when ethnos in ('D','E','F','G') then 'Mixed'
                                        when ethnos in ('H','J','K','L','4','5','6') then
'Asian or Asian British'
                                        when ethnos in ('M','N','P','1','2','3') then 'Black
or Black British'
                                        when ethnos in ('R','7') then 'Chinese'
                                        when ethnos in ('S','8') then 'Other Ethnic Groups'
                                        when ethnos in ('Z','X','9') then 'Unknown' end
----158642


alter table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add  Charlson smallint;

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Charlson = t2.DOMI_0214_CHARLSON_NWT
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_SPELLS t2
                                        on t1.PAtient_ID = t2.Patient_ID
                                        and t1.Spell_ID = t2.Spell_ID
where t2.PT_SpellYear in (2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014)
        ---(158642 row(s) affected)


alter table [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add first_surgery datetime,
first_chemo datetime,
first_radiotherapy datetime,
Pre_cancer int

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
OP_DATE

                                from HED_Basetables.dbo.Hes_OPER_Extract

                                where oper_code like 'x6[34578]%'

                                        or oper_code  like 'Y9[12]%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.OP_DATE
        -----    (5818 row(s) affected)



update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
EPISTART
```

```
                                        from HED_Basetables.dbo.HES_DIAG_Extract

                                        where ICD10 in ('z081' , 'z510' ,'z541')) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.EPISTART
----(7086 row(s) affected)


---Updating chemotherapy from outpatients

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200607

                                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373')) t2
                                                                                    on
t1.Patient_ID = t2.PatientID collate database_Default
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
----(518 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200708

                                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373')) t2
                                                                                    on
t1.Patient_ID = t2.PatientID collate database_Default
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
----(611 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200809

                                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373'))  t2
                                                                                    on
t1.Patient_ID = t2.PatientID collate database_Default
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
---     (856 row(s) affected)


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200910

                                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373')) t2
                                                                                    on
t1.Patient_ID = t2.PatientID collate database_Default
```

```
                                                                                and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
----    (1025 row(s) affected)


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201011

                                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373'))t2
                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
-----(1161 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201112

                                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373')) t2
                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
----(1490 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201213

                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373')) t2
                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                and
t1.firstAdmissionDiagnosis <= t2.APPTDATE



----(1011 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201314

                        where opertn_01 like '%x7[0123]%'

                                                or opertn_01 in ( 'X352','X384', 'X373')) t2
                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
        ---816

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
set chemo = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201415

                          where opertn_01 like '%x7[0123]%'

                                              or opertn_01 in ( 'X352','X384', 'X373')) t2
                                                                                          on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                          and
t1.firstAdmissionDiagnosis <= t2.APPTDATE

--574
---and updateing radiotherapy again (from outpatients)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
APPTDATE

                                        from HES_DW.dbo.HES_OP_200607

                                        where opertn_01  like  'x6[34578]%'      or opertn_01 like
'Y9[12]%') t2
                                                                                          on
t1.Patient_ID = t2.PatientID collate database_Default
                                                                                          and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
----(193 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
APPTDATE

                                        from HES_DW.dbo.HES_OP_200708

                                        where opertn_01  like  'x6[34578]%'      or opertn_01  like
'Y9[12]%') t2
                                                                                          on
t1.Patient_ID = t2.PatientID collate database_Default
                                                                                          and
t1.firstAdmissionDiagnosis <= t2.APPTDATE

----    (258 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
APPTDATE

                                        from HES_DW.dbo.HES_OP_200809

                                        where opertn_01  like  'x6[34578]%'      or opertn_01 like
'Y9[12]%') t2
                                                                                          on
t1.Patient_ID = t2.PatientID collate database_Default
                                                                                          and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
----(508 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PatientID,
APPTDATE

                                        from HES_DW.dbo.HES_OP_200910

                                        where opertn_01  like  'x6[34578]%'      or opertn_01  like
'Y9[12]%') t2
                                                                                          on
t1.Patient_ID = t2.PatientID collate database_Default
                                                                                          and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
---(929 row(s) affected)
```

```
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
APPTDATE

                                    from HES_DW.dbo.HES_OP_201011

                                    where opertn_01  like  'x6[34578]%'      or opertn_01  like
'Y9[12]%') t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
----(1026 row(s) affected)

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
APPTDATE

                                    from HES_DW.dbo.HES_OP_201112

                                    where opertn_01  like  'x6[34578]%'      or opertn_01  like
'Y9[12]%') t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.APPTDATE

----(1165 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
APPTDATE

                                    from HES_DW.dbo.HES_OP_201213

                                    where opertn_01  like  'x6[34578]%'      or opertn_01  like
'Y9[12]%') t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.APPTDATE

---     (1030 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
APPTDATE

                                    from HES_DW.dbo.HES_OP_201314

                                    where opertn_01  like  'x6[34578]%'      or opertn_01  like
'Y9[12]%') t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.APPTDATE

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set radiotherapy = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
APPTDATE

                                    from HES_DW.dbo.HES_OP_201415

                                    where opertn_01  like  'x6[34578]%'      or opertn_01 like
'Y9[12]%') t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.APPTDATE
```

```
---720

SET CONCAT_NULL_YIELDS_NULL OFF
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set first_surgery  = opdate
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
min(op_date) opdate

                                        from (select Patient_ID,  op_date

                                                                        from
HED_Basetables.dbo.HES_OPER_Extract

                                                                        where oper_code  like 'G0[123]%'

                                                                                or oper_code like
'%G2[78]%') a


                                        group by Patient_ID) t2
                                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                                        and
t1.firstAdmissionDiagnosis <= t2.opdate
----    (35646 row(s) affected)

----check this has worked...
select distinct Surgery, first_surgery
from    [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
Order by                first_surgery,          Surgery



---Lets check the data again ...
select  Surgery, first_surgery, count(*)
from    [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group  by               first_surgery,          Surgery
order  by               first_surgery,          Surgery
---this now looks right!!!!


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set pre_cancer = 1
from    [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses       t1 inner join (

                                select PAtient_ID, epistart

                                from HED_Basetables.dbo.HES_Diag_extract

                                where ICD10  like 'C[023456789]%'

                                        or ICD10  like 'C1[01234789]%'

                                ) t2
                                                        on t1.Patient_ID = t2.Patient_ID
                                                        and t1.firstAdmissionDiagnosis > t2.EPISTART

----    (23822 row(s) affected)
/*alter table  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add LSOA varchar(20),
Ward varchar(10),
LA varchar(10); */


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set LSOA  = t2.SOAL
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_EPISODEs t2
                                                                        on t1.Patient_ID =
t2.PATIENT_ID
                                                                        and t1.SPELL_ID =
t2.SPELL_ID
```

```
where PT_SpellYear in (2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014, 2015,2016)
----(158642 row(s) affected)
select LSOA, COUNT(*)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
group by        LSOA
order by LSOA
----566 missing
select LSOA, COUNT(*)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by        LSOA
order by LSOA



----we need to alter the table so we can exclude those patients undergoing Barrets surveillance
/*alter table  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add Barretts int; */



-----118359

update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set barretts = 0 ;
---(195126 row(s) affected)

update  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set barretts = 1
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join     (select
t2.Patient_ID

                                    from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join  (select patient_ID, epistart


        from HED_basetables.dbo.HES_Diag_extract


        where  icd10 like  'k227%') t2

                                                                        on t1.Patient_ID =
t2.Patient_ID

                                                                        and datediff(mm,
endoscopy_date , t2.epistart)  < 30) t2
                                                                                        on
t1.Patient_ID = t2.PAtient_ID
----4572




-----we want to add something now for alarm signals
alter table  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add anaemia int,
abdominalmass int,
Dysphagia int,
Dyspepsia int,
Haematemesis int,
Melaena int,
Vomiting int,
Weightloss int,
AbnormalRadiologyImaging int,
Cancer_code varchar(4),
site_cancer varchar(10),
biopsy int,
updated_cancer_code varchar(4),
Gastric_polyp int,
oesophageal_polyp int,
d2_biop int,
```

```
Benign_path int,
GastricUlcerBiopsy int,
Osophagitis int,
Duodenal_ulcer int,
duodenal_polyp int       ;


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set anaemia  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set abdominalmass  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Dysphagia  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Dyspepsia  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Haematemesis  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Melaena  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Vomiting  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Weightloss  = 0 ;
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set AbnormalRadiologyImaging  = 0 ;


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set anaemia  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%D50[89]%' ;
----(6634 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set abdominalmass  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%R190%' ;
----(368 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Dysphagia  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%R13%' ;
---7715

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Dyspepsia  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%K30%'
        and STARTAGE > 55 ;
----(2270 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Haematemesis  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%K920%' ;
----(2460 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Melaena  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%K921%' ;
----(2614 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Vomiting  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%R11%' ;
----(2144 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set Weightloss  = 1
```

```
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%R634%' ;
----(4907 row(s) affected)
update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set AbnormalRadiologyImaging  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%R93[23]%' ;
---(580 row(s) affected)

---and to start updatin

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cancer_code = t2.cancer_code
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join  (select
patient_ID,epistart, ICD10 cancer_code

                                                    from  HED_Basetables.dbo.HES_DIAG_Extract

                                                    where ICD10 like 'C1[56]%'       ) t2
                                                                on t1.Patient_ID =
t2.Patient_ID
                                                                and
t1.firstAdmissionDiagnosis       = t2.EPISTART
---             (152379 row(s) affected)

----LEts check this has worked
select  cancer_code, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by cancer_code


Update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cancer_code = Underlying
where cancer_code is null and underlying like 'C1[56]%'

Update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cancer_code = cause_3
where cancer_code is null and cause_3 like 'C1[56]%'


update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cancer_code = t2.cancer_code
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join  (select
patient_ID,epistart, ICD10 cancer_code

                                                    from  HED_Basetables.dbo.HES_DIAG_Extract

                                                    where ICD10 like 'C1[56]%'       ) t2
                                                                on t1.Patient_ID =
t2.Patient_ID
                                                                and datediff(dd,
t1.firstAdmissionDiagnosis       ,  t2.EPISTART   ) < 30 and  t1.cancer_code is null

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cancer_code = t2.cancer_code
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join  (select
patient_ID,epistart, ICD10 cancer_code

                                                    from  HED_Basetables.dbo.HES_DIAG_Extract

                                                    where ICD10 like 'C1[56]%'       ) t2
                                                                on t1.Patient_ID =
t2.Patient_ID
                                                                and datediff(dd,
t1.firstAdmissionDiagnosis       ,  t2.EPISTART   ) < 90 and  t1.cancer_code is null
---Lets check the records now
select  cancer_code, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by cancer_code
---6 nulls how frustrating
```

```
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Cancer_code = left(t2.DIAG_01,4)
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join  #UpperGIDiagnosis_OP t2
                                                                        on t1.Patient_ID =
t2.PatientID
                                                                        and datediff(dd,
t1.firstAdmissionDiagnosis       ,  t2.minadmission       ) < 90 and  t1.cancer_code is null

select Patient_ID
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and cancer_code is null
---I have nine IDs here - some are duplicate though
select Patient_ID
into #t6
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
group by Patient_ID
having  COUNT(distinct cancer_code) =0
---no they all have one record and there are only six without one


select distinct  a.Patient_ID, cast(a.ICD10 as varchar) ICD10, a.SPELL_ID
into #t1a
from hed_basetables.dbo.HES_DIAG_Extract a inner join #t1 b
                        on a.Patient_ID = b.PAtient_ID
where left(ICD10, 3) in ('c15','c16')


select PatientID, Apptdate, diag_01
into #t1b
        from HES_DW.dbo.HES_OP_200607
        where  diag_01 like 'c1[56]%'
        union All
        select PatientID , Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_200708
        where diag_01 like 'c1[56]%'
                union All
        select PatientID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_200809
        where diag_01 like 'c1[56]%'
                union All
        select PatientID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_200910
        where diag_01 like 'c1[56]%'
                union All
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201011
        where diag_01 like 'c1[56]%'
                union All
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201112
        where diag_01 like 'c1[56]%'
                        union All
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201213
        where diag_01 like 'c1[56]%'

        union All
        select Patient_ID, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201314
        where diag_01 like 'c1[56]%'
        union All
        select Patient_ID collate database_default, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201415
        where diag_01 like 'c1[56]%'
        union all
                select Patient_ID collate database_default, Apptdate, diag_01 collate database_default
        from HES_DW.dbo.HES_OP_201516
        where diag_01 like 'c1[56]%'

select Patient_ID, ICD10, count(distinct spell_ID ) spells
into #t1d
from (
select PAtient_ID, icd10, cast(spell_ID as varchar) spell_ID
```

```
from #t1a
union all
select PAtientID, diag_01 , cast(apptdate as varchar)
from #t1b ) a
group by Patient_ID, ICD10



select Patient_ID, count(distinct ICD10)
from #t1d
where right(left(ICD10,4),1) <> '9'
group by Patient_ID
order by count(distinct ICD10) desc
---4087 > 1


update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set updated_cancer_code = c.icd10
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses a inner join (select Patient_ID
                                                    from #t1d
                                                    where right(left(ICD10,4),1) <> '9'
                                                    group by Patient_ID
                                                    having count(distinct ICD10) = 1) b
                                on a.Patient_ID = b.Patient_ID
                                                                                    inner join
(select Patient_ID       , left(icd10,4) icd10
                from #t1d
                where right(left(ICD10,4),1) <> '9') c
                                                                                    on
a.Patient_ID = c.Patient_ID
----26821
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set updated_cancer_code = cancer_code
where updated_cancer_code is null
----26821
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set updated_cancer_code = c.icd10
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses a inner join (select Patient_ID,
max(spells) spells
                                                    from #t1d
                                                    where right(left(ICD10,4),1) <> '9'
                                                    group by Patient_ID
                                                    having count(distinct ICD10) = 1) b
                                on a.Patient_ID = b.Patient_ID
                                                                                    inner join
(select Patient_ID,      spells, left(icd10,4) icd10
                from #t1d
                where right(left(ICD10,4),1) <> '9') c
                                                                                    on
b.Patient_ID = c.Patient_ID
                                                                                    and b.spells
= c.spells
select  cancer_code, updated_cancer_code, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by cancer_code     , updated_cancer_code
order by Cancer_code
```

```
Update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set updated_cancer_code = case when left(underlying,3) in ('C15','C16')  and updated_cancer_code is
null then Underlying
                                                                when  left(cause_1,3) in
('C15','C16') and updated_cancer_code is null then cause_1
                                                                when  left(cause_2,3) in
('C15','C16') and updated_cancer_code is null then cause_2
                                                                when  left(cause_3,3) in
('C15','C16') and updated_cancer_code is null then cause_3  else updated_cancer_code end
where updated_cancer_code is null

select Patient_ID, max(case when left(b.CAUSE_OF_DEATH ,3) in ('C15','C16') then b.CAUSE_OF_DEATH
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_1 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_1
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_2 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_2
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_3 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_3
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_4 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_4
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_5 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_5
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_6 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_6
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_7 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_7
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_8 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_8
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_9 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_9
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_10 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_10
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_11 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_11
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_12 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_12
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_13 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_13
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_14 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_14
                                                when left(b.CAUSE_OF_DEATH_NON_NEONATAL_15 ,3) in
('C15','C16') then b.CAUSE_OF_DEATH_NON_NEONATAL_15 end) icd10
into #t1e
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses     a inner join
HES_DW_ONS.dbo.HED_ONS_Mortality b
                                on a.patient_ID = b.EXTRACT_HESID
                                and a.updated_cancer_code is null
group by patient_ID


update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set updated_cancer_code = icd10
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses a inner join #t1e b
                                on a.Patient_ID = b.Patient_ID
---as exprected the largest number of codes are in the '9' category
update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set site_cancer = case when left(updated_cancer_code,3) = 'C15' then 'Oesophag' else 'Gastric' end
---(195126 row(s) affected)
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set biopsy = case when
OPERTN_01+OPERTN_02+OPERTN_03+OPERTN_04+OPERTN_05+OPERTN_06+OPERTN_07+opertn_08+opertn_09+OPERTN_10+OP
ERTN_11+opertn_12 like '%G161%' then 1
                                 when
OPERTN_01+OPERTN_02+OPERTN_03+OPERTN_04+OPERTN_05+OPERTN_06+OPERTN_07+opertn_08+opertn_09+OPERTN_10+OP
ERTN_11+opertn_12 like '%G451%' then 1 else 0 end
----(195126 row(s) affected)


alter table Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
add Diag_date_altered datetime;
--and now to update this
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diag_date_altered = t2.Minadate
```

```
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
PAtient_ID, firstAdmissionDiagnosis, Max(endoscopy_date)Minadate,droprecord

                                                from
Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

                                                where dayspostopEndscopyDiag < 183 and droprecord  is
null

                                                group by  PAtient_ID,
firstAdmissionDiagnosis,droprecord

                                                )t2
                                        on t1.Patient_ID = t2.Patient_ID
----(177085 row(s) affected)

---lets look at these records and see whether a) how many patietns have a different first admission
date

select distinct PAtient_ID, firstAdmissionDiagnosis, diag_date_altered, max(endoscopy_date) end_date
into #Endoscopy_date_diags
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
group by PAtient_ID, firstAdmissionDiagnosis, diag_date_altered
----(133507 row(s) affected)
---Lets double check how many patietns there actually are without a colonoscopy

select *
from #Endoscopy_date_diags
where Diag_date_altered is null
---10392

---Right it looks as though there are some patients who had operations in the right area at the
---time of diagnoisis it may be worthwhile updating these records accordingly to the date of first "G"
 ---procedure - this will mean undoing what I have already doen

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diag_date_altered = null
----(195126 row(s) affected)
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diag_date_altered = t2.Minadate
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select distinct
PAtient_ID, firstAdmissionDiagnosis, Max(endoscopy_date)Minadate

                                                from
Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

                                                where dayspostopEndscopyDiag < 5

                                                group by  PAtient_ID, firstAdmissionDiagnosis)t2
                                        on t1.Patient_ID = t2.Patient_ID
----(151978 row(s) affected)
---again lets look at the paients who don't have a record of colonoscopy
select distinct Patient_ID  , firstAdmissionDiagnosis
into #missendoscopy
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where Diag_date_altered is null
----(24330 row(s) affected)

-----this looks interesting and confirms that paitnets seem to be having other kinds of scopes or
hemicolectmies etc
select t1.PAtient_ID, op_date  colonoscopy_date
into  #updated_endoscopydate
from HED_Basetables.dbo.HES_OPER_Extract t1 inner join #missendoscopy t2
                                                on t1.Patient_ID = t2.Patient_ID collate
database_default
                                                and abs(datediff(dd, t1.op_date,
t2.firstAdmissionDiagnosis))< 10
 where left(oper_code ,3)  in ('G01', 'G02', 'G03','G04', 'G05', 'G07', 'G10', 'G12','G14',
'G15','G16','G17','G18','G19','G21','G27', 'G28', 'G29', 'G43', 'G44', 'G45', 'G46', 'G49','G50')

---(11202 row(s) affected)

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diag_date_altered = t2.colonoscopy_date
```

```
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join #updated_endoscopydate t2
                                                                                                    on
t1.Patient_id = t2.patient_ID collate database_default
where Diag_date_altered is null
----(12749 row(s) affected)

---Lets have a look to see how many are left without adjusted dates
select distinct Patient_ID
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where Diag_date_altered is null
----22755

select distinct Patient_ID  , firstAdmissionDiagnosis
into #missendoscopy_2
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where Diag_date_altered is null
----18286
---Lets now look to see what is going on with these patietnds

select t1.Patient_ID, t1.ADMIDATE, t1.EPISTART, t2.firstAdmissionDiagnosis, t1.OPERTN_01,
t1.OPERTN_02, t1.OPERTN_03, t1.OPERTN_04, t1.OPERTN_05,
        t1.OPERTN_06, t1.OPERTN_07, t1.OPERTN_08, t1.OPERTN_09, t1.OPERTN_10, t1.OPERTN_11,
t1.opertn_12
from HED_Basetables.dbo.HES_EPISODES t1 inner join #missendoscopy_2 t2
                                                    on t1.Patient_ID = t2.Patient_ID
                                                    and t1.EPISTART <= t2.firstAdmissionDiagnosis
where PT_SpellYear in (2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2001)
order by t1.patient_ID, admidate
-----97816
select t1.PAtient_ID,max( op_date) colonoscopy_date
into #updated_endoscopydate_P2
from  HED_Basetables.dbo.HES_OPER_Extract t1 inner join #missendoscopy_2 t2
                                            on t1.Patient_ID = t2.Patient_ID collate
database_default
                                            and datediff(dd, t1.op_date ,
t2.firstAdmissionDiagnosis) between 1 and 90
where left(oper_code ,3)  in ('G01', 'G02', 'G03','G04', 'G05', 'G07', 'G10', 'G12','G14',
'G15','G16','G17','G18','G19','G21','G27', 'G28', 'G29', 'G43', 'G44', 'G45', 'G46', 'G49','G50')
group by t1.Patient_ID
---there are a few updateable fields here
---             9324 row(s) affected)

---(7213 row(s) affected)
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diag_date_altered = t2.colonoscopy_date
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join #updated_endoscopydate_P2
t2
                                                                                                    on
t1.Patient_id = t2.patient_ID collate database_default
where Diag_date_altered is null
----(18232 row(s) affected)


---lets see whether this has improved things
select distinct Patient_ID
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where Diag_date_altered is null
----10338 rows - so we are definitely gewtting there
---I think the best thing to do for these patients is to set the diagnosis date to that of the first
admission date

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Diag_date_altered = firstAdmissionDiagnosis
where Diag_date_altered is null
---(18384 row(s) affected)

---Lets put in the altered tiem from endoscopy
alter table Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
add dayspostop_EndoscpyDiag_altered int;

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set dayspostop_EndoscpyDiag_altered = DATEDIFF(dd, endoscopy_date, case when Diag_date_altered IS null
then firstAdmissionDiagnosis else Diag_date_altered end)
-----(195126 row(s) affected)
```

```
---I'll ened to go through and alter the droprecords now thoguh

---I'm going to have to add inofmration for the following columns  Gastric_polyp, oesophageal_polyp,
d2_biop, benign_path, GastricUlcerBiopsy, Osophagitis, Duodenal_ulcer, Duodenal_polyp
set concat_null_yields_null off
select t1.Patient_ID
from HED_basetables.dbo.HES_DIAG_Extract t1  inner join
Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t2
                                on t1.PAtient_ID = t2.PAtient_ID
                                and t1.epistart <= t2.firstadmissiondiagnosis
where icd10 in ('K317','D131','D132')
group by t1.patient_ID



---I'll also have to add in Mets information
alter table Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
add liver_mets tinyint,
LLung_mets tinyint,
Bone_mets tinyint,
retroperti_perit_mets tinyint;

--Lets update these records
set concat_null_yields_null off
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LLung_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10 = 'C780'
 -----(6007 row(s) affected)
---Lets look at adding in if the patient has primary lung cancer records (this is probably due to
miscoding)
set concat_null_yields_null off
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LLung_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10 like  'C34%'
 -----(1496 row(s) affected)


update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set retroperti_perit_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10 = 'C786'

 -----(5893 row(s) affected)
---Now onto the liver I've seen some evidence about 25% of cancers have liver mets
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Liver_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10 =  'C787'
 -----(13187 row(s) affected)

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Bone_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10 = 'C795'
 -----(3883 row(s) affected)

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
set Bone_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10  like 'c4[01]%'

---(20 row(s) affected)


update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set retroperti_perit_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10   like 'c48%'

----(83 row(s) affected)
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Liver_mets = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10 like 'c22%'
----(264 row(s) affected)



-----I also need to create a variable for looking at lymph node mets
--alter table Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
---add lymphNodeMet int

update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set lymphNodeMet = 1
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HED_Basetables.dbo.HES_diag_extract t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.Diag_date_altered, t2.EPISTART) between 0 and 365
where icd10  like 'C77[1245]%'
 -----(12911 row(s) affected)




update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Liver_Mets = 1
where Patient_ID  in (select distinct t1.Patient_ID
                                                        from
        Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2

                                            on t1.Patient_ID = t2.EXTRACT_HESID collate
database_default
                                                        where droprecord is null and
Liver_mets is null and DATEDIFF(dd, Diag_date_altered, t2.DOD) between 0 and 365 and (underlying in(
'C797', 'D015', 'D376') or cause_1 in( 'C797', 'D015', 'D376')
                                                                or cause_2 in(
'C797', 'D015', 'D376') or cause_3 in( 'C797', 'D015', 'D376') or CAUSE_OF_DEATH_NON_NEONATAL_4 in(
'C797', 'D015', 'D376')or CAUSE_OF_DEATH_NON_NEONATAL_5 in( 'C797', 'D015', 'D376')or
CAUSE_OF_DEATH_NON_NEONATAL_6 in( 'C797', 'D015', 'D376')
                                                                or
CAUSE_OF_DEATH_NON_NEONATAL_7 in( 'C797', 'D015', 'D376')or CAUSE_OF_DEATH_NON_NEONATAL_8 in( 'C797',
'D015', 'D376')or CAUSE_OF_DEATH_NON_NEONATAL_9 in( 'C797', 'D015', 'D376') or
CAUSE_OF_DEATH_NON_NEONATAL_10 in( 'C797', 'D015', 'D376')
                                                                or
CAUSE_OF_DEATH_NON_NEONATAL_11 in( 'C797', 'D015', 'D376') or CAUSE_OF_DEATH_NON_NEONATAL_12 in(
'C797', 'D015', 'D376')
                                                                or      underlying
like'C22%' or cause_1 like 'C22%'        or cause_2 like 'C22%' or cause_3  like 'C22%' or
CAUSE_OF_DEATH_NON_NEONATAL_4 like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_5 like 'C22%' or
CAUSE_OF_DEATH_NON_NEONATAL_6  like 'C22%'
```

```
                                                                                    or
CAUSE_OF_DEATH_NON_NEONATAL_7  like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_8  like 'C22%' or
CAUSE_OF_DEATH_NON_NEONATAL_9  like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_10  like 'C22%'
                                                                                    or
CAUSE_OF_DEATH_NON_NEONATAL_11  like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_12  like 'C22%'))
-----(121 row(s) affected)

----12 rows
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID  in (select distinct Patient_ID
from    Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                                    on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
where droprecord is null and lymphNodeMet is null and DATEDIFF(dd, Diag_date_altered, t2.DOD) between
0 and 365 and (underlying like '%C77[1245]%' or cause_1  like '%C77[1245]%'
               or cause_2  like '%C77[1245]%' or cause_3  like '%C77[1245]%' or
CAUSE_OF_DEATH_NON_NEONATAL_4  like '%C77[1245]%' or CAUSE_OF_DEATH_NON_NEONATAL_5 like
'%C77[1245]%'or CAUSE_OF_DEATH_NON_NEONATAL_6  like '%C77[1245]%'
               or CAUSE_OF_DEATH_NON_NEONATAL_7  like '%C77[1245]%'or CAUSE_OF_DEATH_NON_NEONATAL_8
like '%C77[1245]%' or CAUSE_OF_DEATH_NON_NEONATAL_9  like '%C77[1245]%'or
CAUSE_OF_DEATH_NON_NEONATAL_10  like '%C77[1245]%'
               or CAUSE_OF_DEATH_NON_NEONATAL_11  like '%C77[1245]%' or
CAUSE_OF_DEATH_NON_NEONATAL_12  like '%C77[1245]%'
               ))
----(15 row(s) affected)
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID  in (select distinct Patient_ID
                                                    from
        Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2

                                on t1.Patient_ID = t2.EXTRACT_HESID collate database_default
                                                    where droprecord is null and Bone_mets is null and
DATEDIFF(dd, Diag_date_altered, t2.DOD) between 0 and 365 and (underlying like '%C795%' or cause_1
like '%C795%'
                                                                    or cause_2  like '%C795%' or cause_3
like '%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_4  like '%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_5 like
'%C795%'or CAUSE_OF_DEATH_NON_NEONATAL_6  like '%C795%'
                                                                    or CAUSE_OF_DEATH_NON_NEONATAL_7
like '%C795%'or CAUSE_OF_DEATH_NON_NEONATAL_8  like '%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_9  like
'%C795%'or CAUSE_OF_DEATH_NON_NEONATAL_10  like '%C795%'
                                                                    or CAUSE_OF_DEATH_NON_NEONATAL_11
like '%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_12  like '%C795%'
                                                                    ))
----(143 row(s) affected)
---124
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID  in (select distinct Patient_ID
                                                                    from
        Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                                    on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                                    where droprecord is null and
retroperti_perit_mets is null and DATEDIFF(dd, Diag_date_altered, t2.DOD) between 0 and 365 and
(underlying like '%C786%' or cause_1  like '%C786%'
                                                                                    or cause_2  like
'%C786%' or cause_3  like '%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_4  like '%C786%' or
CAUSE_OF_DEATH_NON_NEONATAL_5 like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_6  like '%C786%'
                                                                                    or
CAUSE_OF_DEATH_NON_NEONATAL_7  like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_8  like '%C786%' or
CAUSE_OF_DEATH_NON_NEONATAL_9  like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_10  like '%C786%'
                                                                                    or
CAUSE_OF_DEATH_NON_NEONATAL_11  like '%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_12  like '%C786%'
                                                                                    ))
----(156 row(s) affected)NEONATAL_11 in( 'C780', 'C34') or CAUSE_OF_DEATH_NON_NEONATAL_12 in( 'C780',
'C34')

----335
```

```
---Need to add day of endoscopy , cancer location, and update gastric polyp, oesophageal polyp, d2
biopsy, benign pathology, gastric ulcer biopsy, oesophagitis, duodenal ulcer, duodenal polyp
---emergency procedures - another file
--adding in day of endoscopy, cancer location

/*alter table Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
add day_scope varchar(10),
cancer_location varchar(20); */


update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set day_scope = DATENAME(dw, endoscopy_date)
----(195126 row(s) affected)


update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set cancer_location = case when updated_cancer_code in ('c150','C153','c151','C154') then 'Proximal
Oesophagus'
                                                    when  updated_cancer_code in ('C152','C155')
then 'Distal Oesophagus'
                                                    when  updated_cancer_code in ('c158','C159')
then 'Oesophagus_NOS'
                                                    when updated_cancer_code in ('c160','C161')
then 'Proximal_stomach'
                                                    when  updated_cancer_code in ('c162','C165',
'C166') then 'Mid_stomach'
                                                    when  updated_cancer_code in ('c163','C164')
then 'Distal_Stomach '
                                                    when  updated_cancer_code in ('c168','C169')
then 'Stomach_NOS' end
----(195126 row(s) affected)

select top 100 *
from Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

---and now to update the pathology columns
update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set gastric_polyp = case when diag_01 in ('K137','D131') then 1
                                                when diag_02 in ('K137','D131') then 1
                                                when diag_03 in ('K137','D131') then 1
                                                when diag_04 in ('K137','D131') then 1
                                                when diag_05 in ('K137','D131') then 1
                                                when diag_06 in ('K137','D131') then 1
                                                when diag_07 in ('K137','D131') then 1
                                                when diag_08 in ('K137','D131') then 1
                                                when diag_09 in ('K137','D131') then 1
                                                when diag_10 in ('K137','D131') then 1
                                                when diag_11 in ('K137','D131') then 1
                                                when diag_12 in ('K137','D131') then 1
                                                when diag_13 in ('K137','D131') then 1
                                                when diag_14 in ('K137','D131') then 1 else 0 end
---(195126 row(s) affected)
update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set oesophageal_polyp = case when diag_01 in ('K228','D130') then 1
                                                when diag_02 in ('K228','D130') then 1
                                                when diag_03 in ('K228','D130') then 1
                                                when diag_04 in ('K228','D130') then 1
                                                when diag_05 in ('K228','D130') then 1
                                                when diag_06 in ('K228','D130') then 1
                                                when diag_07 in ('K228','D130') then 1
                                                when diag_08 in ('K228','D130') then 1
                                                when diag_09 in ('K228','D130') then 1
                                                when diag_10 in ('K228','D130') then 1
                                                when diag_11 in ('K228','D130') then 1
                                                when diag_12 in ('K228','D130') then 1
                                                when diag_13 in ('K228','D130') then 1
                                                when diag_14 in ('K228','D130') then 1 else 0 end
---(195126 row(s) affected)
update Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set d2_biop = case when
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12+opertn_13+opertn_14 like '%G451Z274%' then 1 else 0 end
---(195126 row(s) affected)
update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
set benign_path = case when diag_01 in ('D130', 'D131', 'D132') then 1
                                                when diag_02 in ('D130', 'D131', 'D132') then 1
                                                when diag_03 in ('D130', 'D131', 'D132') then 1
                                                when diag_04 in ('D130', 'D131', 'D132') then 1
                                                when diag_05 in ('D130', 'D131', 'D132') then 1
                                                when diag_06 in ('D130', 'D131', 'D132') then 1
                                                when diag_07 in ('D130', 'D131', 'D132') then 1
                                                when diag_08 in ('D130', 'D131', 'D132') then 1
                                                when diag_09 in ('D130', 'D131', 'D132') then 1
                                                when diag_10 in ('D130', 'D131', 'D132') then 1
                                                when diag_11 in ('D130', 'D131', 'D132') then 1
                                                when diag_12 in ('D130', 'D131', 'D132') then 1
                                                when diag_13 in ('D130', 'D131', 'D132') then 1
                                                when diag_14 in ('D130', 'D131', 'D132') then 1 else
0 end

update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set GastricUlcerBiopsy = case when diag_01 like 'K25%' then 1
                                                when diag_02 like 'K25%' then 1
                                                when diag_03 like 'K25%' then 1
                                                when diag_04 like 'K25%' then 1
                                                when diag_05 like 'K25%' then 1
                                                when diag_06 like 'K25%' then 1
                                                when diag_07 like 'K25%' then 1
                                                when diag_08 like 'K25%' then 1
                                                when diag_09 like 'K25%' then 1
                                                when diag_10 like 'K25%' then 1
                                                when diag_11 like 'K25%' then 1
                                                when diag_12 like 'K25%' then 1
                                                when diag_13 like 'K25%' then 1
                                                when diag_14 like 'K25%' then 1 else 0 end
---(195126 row(s) affected)

update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Osophagitis = case when diag_01 like 'K20%' then 1
                                                when diag_02 like 'K20%' then 1
                                                when diag_03 like 'K20%' then 1
                                                when diag_04 like 'K20%' then 1
                                                when diag_05 like 'K20%' then 1
                                                when diag_06 like 'K20%' then 1
                                                when diag_07 like 'K20%' then 1
                                                when diag_08 like 'K20%' then 1
                                                when diag_09 like 'K20%' then 1
                                                when diag_10 like 'K20%' then 1
                                                when diag_11 like 'K20%' then 1
                                                when diag_12 like 'K20%' then 1
                                                when diag_13 like 'K20%' then 1
                                                when diag_14 like 'K20%' then 1 else 0 end
---(195126 row(s) affected)
update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Duodenal_ulcer = case when diag_01 like 'K26%' then 1
                                                when diag_02 like 'K26%' then 1
                                                when diag_03 like 'K26%' then 1
                                                when diag_04 like 'K26%' then 1
                                                when diag_05 like 'K26%' then 1
                                                when diag_06 like 'K26%' then 1
                                                when diag_07 like 'K26%' then 1
                                                when diag_08 like 'K26%' then 1
                                                when diag_09 like 'K26%' then 1
                                                when diag_10 like 'K26%' then 1
                                                when diag_11 like 'K26%' then 1
                                                when diag_12 like 'K26%' then 1
                                                when diag_13 like 'K26%' then 1
                                                when diag_14 like 'K26%' then 1 else 0 end
---(195126 row(s) affected)
update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set duodenal_polyp = case when diag_01 like 'D132' then 1
                                                when diag_02 like 'D132' then 1
                                                when diag_03 like 'D132' then 1
                                                when diag_04 like 'D132' then 1
                                                when diag_05 like 'D132' then 1
                                                when diag_06 like 'D132' then 1
                                                when diag_07 like 'D132' then 1
                                                when diag_08 like 'D132' then 1
```

```
                                                when diag_09 like 'D132' then 1
                                                when diag_10 like 'D132' then 1
                                                when diag_11 like 'D132' then 1
                                                when diag_12 like 'D132' then 1
                                                when diag_13 like 'D132' then 1
                                                when diag_14 like 'D132' then 1 else 0 end
---(195126 row(s) affected)

alter table Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
add Oesophageal_ulcer tinyint,
Oesophageal_stricture tinyint,
procode3_adjusted varchar(5),
pepticUlcer tinyint,
GastrojejunalUcler tinyint;

update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set procode3_adjusted = t2.procode3_adjusted
from  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join
cancer_research.dbo.ltblFE_20140109_ProcodeMapped t2
                                                on t1.procode = t2.procode collate database_default
-----           (195126 row(s) affected)


update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set pepticUlcer= case when diag_01 like 'K27%' then 1
                                                when diag_02 like 'K27%' then 1
                                                when diag_03 like 'K27%' then 1
                                                when diag_04 like 'K27%' then 1
                                                when diag_05 like 'K27%' then 1
                                                when diag_06 like 'K27%' then 1
                                                when diag_07 like 'K27%' then 1
                                                when diag_08 like 'K27%' then 1
                                                when diag_09 like 'K27%' then 1
                                                when diag_10 like 'K27%' then 1
                                                when diag_11 like 'K27%' then 1
                                                when diag_12 like 'K27%' then 1
                                                when diag_13 like 'K27%' then 1
                                                when diag_14 like 'K27%' then 1 else 0 end

update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Oesophageal_ulcer = case when diag_01 = 'K221' then 1
                                                when diag_02  = 'K221'  then 1
                                                when diag_03 = 'K221'  then 1
                                                when diag_04  = 'K221'  then 1
                                                when diag_05 = 'K221'  then 1
                                                when diag_06  = 'K221'  then 1
                                                when diag_07  = 'K221' then 1
                                                when diag_08  = 'K221'  then 1
                                                when diag_09  = 'K221'  then 1
                                                when diag_10  = 'K221'  then 1
                                                when diag_11  = 'K221'  then 1
                                                when diag_12  = 'K221'  then 1
                                                when diag_13  = 'K221'  then 1
                                                when diag_14  = 'K221'  then 1 else 0 end


update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set Oesophageal_stricture = case when diag_01 = 'K222' then 1
                                                when diag_02  = 'K222'  then 1
                                                when diag_03 = 'K222'  then 1
                                                when diag_04  = 'K222'  then 1
                                                when diag_05 = 'K222'  then 1
                                                when diag_06  = 'K222'  then 1
                                                when diag_07  = 'K222' then 1
                                                when diag_08  = 'K222'  then 1
                                                when diag_09  = 'K222'  then 1
                                                when diag_10  = 'K222'  then 1
                                                when diag_11  = 'K222'  then 1
                                                when diag_12  = 'K222'  then 1
                                                when diag_13  = 'K222'  then 1
                                                when diag_14  = 'K222'  then 1 else 0 end

update  Cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set GastrojejunalUcler = case when diag_01 like 'K28%' then 1
```

```
                                        when diag_02 like 'K28%' then 1
                                        when diag_03 like 'K28%' then 1
                                        when diag_04 like 'K28%' then 1
                                        when diag_05 like 'K28%' then 1
                                        when diag_06 like 'K28%' then 1
                                        when diag_07 like 'K28%' then 1
                                        when diag_08 like 'K28%' then 1
                                        when diag_09 like 'K28%' then 1
                                        when diag_10 like 'K28%' then 1
                                        when diag_11 like 'K28%' then 1
                                        when diag_12 like 'K28%' then 1
                                        when diag_13 like 'K28%' then 1
                                        when diag_14 like 'K28%' then 1 else 0 end


---The way in which last endoscopy is defined has been changed. This script takes account these
chnages.
select Patient_ID, COUNT(*) ops
into #t1
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by patient_ID
Order by COUNT(*) desc
---33, 843 patients with one endoscpy (out of 132077)


select Patient_ID, MIN(admidate) admidate
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and Patient_ID in (select Patient_ID
                                                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                                                    where
droprecord is null
                                                                                    group by
patient_ID
                                                                                    having
COUNT(*) > 1)
        and dayspostop_EndoscpyDiag_altered between 180 and 365
group by Patient_ID



update  cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforeDiag = 1
where Patient_ID in (    select Patient_ID
                                        from #t1
                                        where ops = 1)
----(95031 row(s) affected)
update  cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforeDiag = 1
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join

                                                (select Patient_ID, max(endoscopy_date)
epistart
                                                from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                where Patient_ID in (select Patient_ID
                                                                                        from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

        where droprecord is null

        group by patient_ID

        having COUNT(*) > 1 and MAX(dayspostop_EndoscpyDiag_altered) <= 180 ) and droprecord is null
                                                group by Patient_ID
```

```
                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.endoscopy_date = t2.epistart
-----    (24142 row(s) affected)
update  cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforeDiag = 1
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(endoscopy_date) epistart

                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                        from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1) and droprecord is null

                                                    group by Patient_ID

                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.endoscopy_date = t2.epistart
where t1.dayspostop_EndoscpyDiag_altered > 1826
-----(1235 row(s) affected)
/*
select  PAtient_ID, COUNT(*)
from    cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and LastEndoscopybeforeDiag = 1
group by Patient_ID
order by  COUNT(*) desc
---355 multiple rows
select *
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and LastEndoscopybeforeDiag = 1 and Patient_ID in (select  PAtient_ID

                                                    from
        cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

                                                    where droprecord is null and
LastEndoscopybeforeDiag = 1

                                                    group by Patient_ID

                                                    having  COUNT(*) > 1)
order by Patient_ID, admidate
---these all seem to be multiple episodes
select  PAtient_ID
from    cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having  COUNT(*) > 1 and MAX(LastEndoscopybeforeDiag) = 0 */
----17003
update  cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforeDiag = 1
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(endoscopy_date) epistart

                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
                                                    where Patient_ID in (select Patient_ID
                                                                                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

        where droprecord is null

        group by patient_ID

        having COUNT(*) > 1)
                                                            and droprecord is null   and
dayspostop_EndoscpyDiag_altered between 181 and 365
                                                    group by Patient_ID

                                                    ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.endoscopy_date = t2.epistart
----(4499 row(s) affected)
update  cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforeDiag = 1
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(endoscopy_date) epistart
                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                    where Patient_ID in (select Patient_ID
                                                                                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

        where droprecord is null

        group by patient_ID

        having COUNT(*) > 1 and MAX(LastEndoscopybeforeDiag) = 0)
                                                            and droprecord is null   and
dayspostop_EndoscpyDiag_altered between 366 and 1096
                                                    group by Patient_ID

                                                    ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.endoscopy_date = t2.epistart
----(6588 row(s) affected)
update  cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforeDiag = 1
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(endoscopy_date) epistart
                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                    where Patient_ID in (select Patient_ID
                                                                                                                    from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses

        where droprecord is null
```

```
        group by patient_ID

        having COUNT(*) > 1 and MAX(LastEndoscopybeforeDiag) = 0)
                                                    and droprecord is null   and
dayspostop_EndoscpyDiag_altered between 1097 and 1826
                                                group by Patient_ID

                                                ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.endoscopy_date = t2.epistart
---3293
/*select Patient_ID, sum(LastEndoscopybeforediag)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(LastEndoscopybeforediag)  desc
---there are 5500 patients with multiple records for LastEndoscopybeforediag
select Patient_ID, SPELL_ID, SPELL_FINYR, admidate, EPIORDER, ADMIDATE, DISDATE,
LastEndoscopybeforeDiag, dayspostop_EndoscpyDiag_altered
from  [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null and patient_ID in (select Patient_ID
                                                                        from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                                        where
droprecord is null
                                                                        group by
Patient_ID
                                                                        having
sum(LastEndoscopybeforediag) > 1)
order by Patient_ID
----a lot of these do seem to be due to multiple episodes
---Lets first check that everyone has a last record
select Patient_ID
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having max(LastEndoscopybeforediag) = 0
---there are 2676 patietns here without one so lets look at these records first
select Patient_ID, SPELL_ID, SPELL_FINYR, admidate, EPIORDER, ADMIDATE, DISDATE,
LastEndoscopybeforeDiag, dayspostop_EndoscpyDiag_altered , row_id
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where PAtient_ID in (select Patient_ID
                                                from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                where droprecord is null
                                                group by Patient_ID
                                                having sum(LastEndoscopybeforediag) >1)
order by Patient_ID, dayspostop_EndoscpyDiag_altered
--this is due to the patients having a record that more thn five years ago and also in the last 6
months*/
update  cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforeDiag = 1
from cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(endoscopy_date) epistart
                                                from
cancer_research.dbo.FE_20160316_EndoscopyMissedDiagnoses
                                                where Patient_ID in (select Patient_ID

        from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses

        where droprecord is null
```

```
        group by Patient_ID

        having max(LastEndoscopybeforediag) = 0  ) and droprecord is null
                                                    group by Patient_ID

                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.endoscopy_date = t2.epistart
----    (2705 row(s) affected)
---and back to the multiple records
/*select Patient_ID, sum(LastEndoscopybeforediag)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(LastEndoscopybeforediag)  desc
---there are 902 patients with multiple records for LastEndoscopybeforediag
----a lot of these do seem to be due to multiple episodes        */

select Patient_Id, spell_Id, max(epiorder) minepi
into #endoscopy_2
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where Patient_ID in (select Patient_ID
                                    from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                    where droprecord is null
                                    group by Patient_ID
                                    having sum(LastEndoscopybeforediag) >1)
        and LastEndoscopybeforediag = 1
        group by  Patient_Id, spell_Id
----1144


update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforediag = 0
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join  #endoscopy_2 t2
                                            on t1.Patient_ID = t2.Patient_ID
                                            and t1.SPELL_ID = t2.SPELL_ID
                                            and t1.EPIORDER <> t2.minepi
-----   9023

select Patient_ID, sum(LastEndoscopybeforediag)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(LastEndoscopybeforediag)  desc
---there are 113  patients with multiple records for Lastcolonoscopybeforediag
---there looks to be some patients with duplicate records
select Patient_ID, spell_ID,epiorder, count(*)
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID, spell_ID,epiorder
order by count(*) desc

/*alter table[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
add row_id int

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set row_id =abs(cast(cast(newid() as varbinary) as int))
where row_id = 0 */

update [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
set LastEndoscopybeforediag = 0
from [Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses t1 inner join (select PAtient_ID,
max(row_id) row_id

                                                                    from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
```

```
                                                                    where droprecord is null and
LastEndoscopybeforeDiag = 1

                                                                    group by Patient_ID)  b
                                                                                          on
t1.Patient_ID = b.Patient_ID
                                                                                          and
t1.row_id = b.row_id

                        inner join (select Patient_ID
                                           from
[Cancer_research].dbo.FE_20160316_EndoscopyMissedDiagnoses
                                           where droprecord is null
                                           group by Patient_ID
                                           having sum(LastEndoscopybeforediag)>1) c
                                                                                  on b.Patient_ID =
c.Patient_ID
where  droprecord is null


---havin a look to see how many patietns have been diagnosed with a gastric ulcer
create table cancer_research.dbo.FE_20160524_gastricUlcer(
 Patient_ID  varchar(50),
spell_ID bigint,
admidate datetime,
FirstAdmissionDiagnosis datetime,
disdate datetime,
OGD_3months tinyint,
Ulcer_code char(4),
procode char(5)
primary key (patient_Id, spell_ID))

insert into  cancer_research.dbo.FE_201062014_gastricUlcer(
Patient_ID, admidate, disdate, procode, spell_ID, Ulcer_code)

select  PAtient_ID ,  min(epistart), max(epiend), procode3,spell_ID, min(t1.ICD10)
from HED_basetables.dbo.HES_DIAG_EXTRACT t1
where ICD10 like 'K2[57]%'
group by PAtient_ID , procode3,spell_ID

-----(626760 row(s) affected)

---identifying first admission with gastric ulcer diag
select Patient_ID, min(admidate) firstOperation
into #firstgastriculcer
from  cancer_research.dbo.FE_201062014_gastricUlcer
group  by patient_ID
---updating this into the main table
---(422893 row(s) affected)
update cancer_research.dbo.FE_201062014_gastricUlcer
set FirstAdmissionDiagnosis = firstOperation
from  cancer_research.dbo.FE_201062014_gastricUlcer t1 inner join #firstgastriculcer t2
                                                         on t1.PAtient_ID = t2.PAtient_ID collate
database_default
----(517966 row(s) affected)
----Identifying patients who have had OGD within first three months following diagnosis
update cancer_research.dbo.FE_201062014_gastricUlcer
set OGD_3months = 1
from  cancer_research.dbo.FE_201062014_gastricUlcer t1 inner join
Cancer_research.dbo.FE_20160524_AllEndoscopies t2
                                                                        on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                        and datediff(dd, t1.admidate
, t2.ADMIDATE) > =0
                                                                        and datediff(dd, t1.disdate
, t2.ADMIDATE) < 90
----(338618 row(s) affected)

select left(procode, 3) provider, 100*cast(sum(case when OGD_3months is null then 0 else 1 end ) as
float)/ cast(count(*) as float)  percUlcerFollowup, cast(count(*) as float)
from  cancer_research.dbo.FE_201062014_gastricUlcer
where FirstAdmissionDiagnosis >= admidate and FirstAdmissionDiagnosis<=admidate
group by left(procode, 3)
```

```
order by 100*cast(sum(case when OGD_3months is null then 0 else 1 end ) as float)/ cast(count(*) as
float)  desc


---this script will be used to identify the providers at which the first admissions with Upper GI
CAncer occurred
---first we need to find all the firsst admissions/ OP appointments for people with Upper GI CAncer
select Patient_ID, epistart, Procode3
into #UpperGIDiagnosis
from HEd_basetables.dbo.HES_diag_extract --- this is a table which lists all diagnoses in a ling
rather than wide format
where ICD10 like 'c1[56]%'
----(464947 row(s) affected)

----identify patietns who die from Upper GI CAncer
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  Extract_HESID, Max(DOD) dod
into #UpperGIDiagnosisfromDeaths
from (select Extract_HESID, DOD
        from HES_DW_ONS.dbo.HED_ONS_Mortality
        where
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON_NEONATA
L_11+CAUSE_OF_DEATH_NON_NEONATAL_12+CAUSE_OF_DEATH_NON_NEONATAL_13
        +CAUSE_OF_DEATH_NON_NEONATAL_14+CAUSE_OF_DEATH_NON_NEONATAL_15 like '%c1[56]%'
        ) t1
group by Extract_HESID
-----(136023 row(s) affected)

---checking to see whether there are any patients who died of lower GI CAncer who weren't admitted to
hosp with it
select *
from  #UpperGIDiagnosisfromDeaths t1  left join #UpperGIDiagnosis t2
                                                        on t1.EXTRACT_HESID = t2.Patient_ID collate
database_default
where t2.Patient_ID is null
---there are 14487 patients here who died without a hospital admission owhere lower gi cancer was
recorded
---lets look at OP data
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  PatientID, procode, MIN(apptdate) minadmission
into #UpperGIDiagnosisOP
from (select PatientID, apptdate, procode
        from HES_DW.dbo.HES_OP_200607
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[56]%'
        union All
        select PatientID, apptdate, procode
        from HES_DW.dbo.HES_OP_200708
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[56]%'
        union All
        select PatientID, apptdate, procode
        from HES_DW.dbo.HES_OP_200809
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[56]%'
        union All
        select PatientID, apptdate, procode
        from HES_DW.dbo.HES_OP_200910
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[56]%'
        union All
        select Patient_ID, apptdate, procode
        from HES_DW.dbo.HES_OP_201011
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[56]%'
        union All
        select Patient_ID, apptdate, procode
        from HES_DW.dbo.HES_OP_201112
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[56]%'  ) t1
group by PAtientID, procode


----(7915 row(s) affected)
```

```
---and getting the tables altogether

select Patient_ID, MIN(minadmission)minadmission
Into #UpperGI
from (select Patient_ID,  epistart minadmission
                from #UpperGIDiagnosis
                union all
                select EXTRACT_HESID collate SQL_Latin1_General_CP1_CI_AS, DOD
                from #UpperGIDiagnosisfromDeaths
                union all
                select PatientID collate SQL_Latin1_General_CP1_CI_AS, minadmission
                from #UpperGIDiagnosisOP) a
group by Patient_ID
----(198498 row(s) affected)

---this will allow us to find diagnosing providers
select t1.PAtient_ID, t1.minadmission, case when t2.patient_ID is not null then t2.PROCODE  collate
database_default
                                                                                                when
t3.PatientID is not null then t3.PROCODE  collate database_default  else null end  provider,
                                                                        case when t2.patient_ID is not null
then 'IP'
                                                                                when t3.PatientID is not
null then  'OP' else null end location ,
                                                                        case when t2.patient_ID is not null
and t2.Admimeth like '2%' then 1
                                                                                else null end
emergency_admission
into #GIdiagnosingadmissions
from  #UpperGI t1 left join  HED_basetables.dbo.HES_Episodes t2
                                                        on t1.patient_ID = t2.Patient_ID collate
database_default
                                                        and t1.minadmission = t2.epistart
                                    left join #UpperGIDiagnosisOP t3
                                                        on t1.Patient_ID = t3.PatientID collate
database_default
                                                        and t1.minadmission = t3.minadmission

order by  case when t2.patient_ID is not null then t2.PROCODE collate database_default
                                                                                                when
t3.PatientID is not null then t3.PROCODE  collate database_default  else null end

/* --- the definition has changed slightly now, we need to also find the endoscopy if a endoscopy has
been performed*/
SET CONCAT_NULL_YIELDS_NULL OFF
select *
into  #Endoscopies
from (select t1.Patient_ID, t1.SPELL_ID, case when t1.admimeth like'2%' then 'Emergency' else null end
emergency_adm,procode,
                                t2.op_date  as colonoscopy_date
        from HED_basetables.dbo.HES_EPISODES t1 inner join HED_basetables.dbo.HES_OPER_EXTRACT t2
                                                                                        on t1.spell_ID =
t2.spell_ID
                                                                                        and t1.EPI_ORDR =
t2.EPI_ORDR
                where t1.ADMIDATE < '2012/01/04' and (oper_code like  '%G1[456]%' or  oper_code like
'%G4[23456]%' or oper_code like '%G5[45]%')
) a
----(6241443 row(s) affected)

---Lets look to see whether any of these endoscopies are diagnosing ones
select distinct t2.PAtient_ID, t2.minadmission, Max(case when left(t1.procode,1) = 'R' then
left(t1.Procode,3)

                        when left(t1.procode,1) = '5' then left(t1.procode,3) else left(t1.Procode,3)
end ),

                         t2.location, t3.colonoscopy_date, max(t1.emergency_adm)
from #endoscopies t1 inner join          #GIdiagnosingadmissions t2
                                on t1.PAtient_ID = t2.PAtient_ID collate database_default
                         inner join (select t1.Patient_ID, max(colonoscopy_date) colonoscopy_date,
minadmission
                                                                                        from  #endoscopies t1 inner
join #GIdiagnosingadmissions t2
```

```
                                                                                                    on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                  where (datediff(dd,
t1.colonoscopy_date, t2.minadmission) between -5 and 180 )
                                                                                  group by t1.patient_ID,
minadmission) t3

                                                  on t2.patient_ID = t3.patient_ID collate
database_default
                                                  and t2.minadmission = t3.minadmission
                                                  and t1.colonoscopy_date = t3.colonoscopy_date

group by t2.PAtient_ID, t2.minadmission,   t2.location, t3.colonoscopy_date
order by t2.patient_ID

---this looks like it is doing what it should be,I tink the best way forward now is to put this into a
table and then link across to the diagnoses again
select distinct t2.PAtient_ID, t2.minadmission, Max(case when left(t1.procode,1) = 'R' then
left(t1.Procode,3)

                        when left(t1.procode,1) = '5' then left(t1.procode,3) else left(t1.Procode,3)
end ) provider,

                         t2.location, t3.colonoscopy_date, max(t1.emergency_adm) emerg_adm
into #endoscopy_diagnoses
from #endoscopies t1 inner join          #GIdiagnosingadmissions t2
                               on t1.PAtient_ID = t2.PAtient_ID collate database_default
                         inner join (select t1.Patient_ID, max(colonoscopy_date) colonoscopy_date,
minadmission
                                                                                  from  #endoscopies t1 inner
join #GIdiagnosingadmissions t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID collate database_Default
                                                                                  where (datediff(dd,
t1.colonoscopy_date, t2.minadmission) between -5 and 180 )
                                                                                  group by t1.patient_ID,
minadmission) t3

                                                  on t2.patient_ID = t3.patient_ID collate
database_Default
                                                  and t2.minadmission = t3.minadmission
                                                  and t1.colonoscopy_date = t3.colonoscopy_date

group by t2.PAtient_ID, t2.minadmission,   t2.location, t3.colonoscopy_date
---(122603 row(s) affected)
----- and now to link across
select distinct t1.patient_ID, case when t2.minadmission is not null then t2.colonoscopy_date else
t1.minadmission end as diag_date,
                 max(case when t2.minadmission is not null then t2.provider else

                        case when left(t1.provider,1) = 'R' then left(t1.provider,3)

                        when left(t1.provider,1) = '5' then left(t1.provider,3) else
left(t1.provider,3) end  end   ) as provider,
                case when t2.minadmission is not null then case when t2.emerg_adm = 'emergency' then 1
else 0 end  else case when t1.emergency_admission = 1 then 1 else 0 end end as emerg_adm
into #finalanalysis_endoscopy
from #GIdiagnosingadmissions  t1 left join #endoscopy_diagnoses t2
                                                  on t1.Patient_ID = t2.Patient_ID
                                                  and t1.minadmission = t2.minadmission

group by t1.patient_ID, case when t2.minadmission is not null then t2.colonoscopy_date else
t1.minadmission end , case when t2.minadmission is not null then case when t2.emerg_adm = 'emergency'
then 1 else 0 end  else case when t1.emergency_admission = 1 then 1 else 0 end end
having max(case when t2.minadmission is not null then t2.provider else

                        case when left(t1.provider,1) = 'R' then left(t1.provider,3)

                        when left(t1.provider,1) = '5' then left(t1.provider,3) else
left(t1.provider,3) end  end   ) is not null
-----(186694 row(s) affected)
--have a look to see how many emergency admissions and patients there are per provider
select provider, sum(emerg_adm), count(distinct patient_ID)
from #finalanalysis_endoscopy
```

```
group by provider

/*
drop table #Endoscopies
drop table #endoscopy_diagnoses
drop table #finalanalysis_endoscopy
drop table #UpperGI
drop table #UpperGIDiagnosis
drop table #UpperGIDiagnosisfromDeaths
drop table #UpperGIDiagnosisOP*/

-- generate a table showing the emergency admissions rate per provider, for the providers with at
least 20 patients diagnosed there
select case when t2.procode3_adjusted is null and left(provider, 3) = 'R1k' then provider collate
database_default else procode3_adjusted end , 100* cast(sum(case when emergency_admission = 1 then 1
else 0 end ) as float)/ cast(count(Patient_ID)as float)
from #GIdiagnosingadmissions t1 left join Cancer_research.dbo.ltblFE_20140109_ProcodeMapped t2
                                                on left(t1.provider,3) = t2.procode collate
database_default
group by case when t2.procode3_adjusted is null and left(provider, 3) = 'R1k' then provider  collate
database_default else procode3_adjusted end
having count(patient_ID) > 20
order by case when t2.procode3_adjusted is null and left(provider, 3) = 'R1k' then provider collate
database_default else procode3_adjusted end
```

Appendix B – SQL codes for PCCRC (chapter 7)

```
----start by identifyinng the colonoscopies
SET CONCAT_NULL_YIELDS_NULL OFF
select Patient_ID, SPELL_ID, SPELL_FINYR, EPIORDER, ADMIDATE, DISDATE, ADMIMETH, CLASSPAT, DIAG_01,
DIAG_02, DIAG_03, DIAG_04,DIAG_05, DIAG_06,
DIAG_07, DIAG_08,DIAG_09, DIAG_10,DIAG_11, DIAG_12, DIAG_13, DIAG_14, DISMETH, GPPRAC, MAINSPEF,
OPERTN_01, OPERTN_02, OPERTN_03, OPERTN_04, OPERTN_05,
OPERTN_06, OPERTN_07, opertn_08, OPERTN_09, OPERTN_10, OPERTN_11, OPERTN_12, OPERTN_13, OPERTN_14,
LEFT(procode,3) procode, SEX, STARTAGE, quintiles , resgor
from HED.dbo.HES_EPISODES
where (opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+ OPERTN_06+opertn_07+ OPERTN_08+
opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+ OPERTN_16+opertn_17+ OPERTN_18+
opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24 like '%H20[16]%' or
                                opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+
OPERTN_06+opertn_07+ OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+
OPERTN_16+opertn_17+ OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24   like
'%H22%')
                                and ADMIDATE < '2012/01/04'
Order by Patient_ID, admidate
-----1800722 endoscopies from 2006/07 onwards
---this is a lot
---looking at the number of colonoscopies between 2001/02 and 06/07
SET CONCAT_NULL_YIELDS_NULL OFF
select Patient_ID, SPELL_ID, SPELL_FINYR, EPIORDER, ADMIDATE, DISDATE, ADMIMETH, CLASSPAT, DIAG_01,
DIAG_02, DIAG_03, DIAG_04,DIAG_05, DIAG_06,
DIAG_07, DIAG_08,DIAG_09, DIAG_10,DIAG_11, DIAG_12, DIAG_13, DIAG_14, DISMETH, GPPRAC, MAINSPEF,
OPERTN_01, OPERTN_02, OPERTN_03, OPERTN_04, OPERTN_05,
OPERTN_06, OPERTN_07, opertn_08, OPERTN_09, OPERTN_10, OPERTN_11, OPERTN_12, OPERTN_13, OPERTN_14,
LEFT(procode,3) procode, SEX, STARTAGE, quintiles , resgor
from HED.dbo.HES_EPISODES_part1
where (opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+ OPERTN_06+opertn_07+ OPERTN_08+
opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+ OPERTN_16+opertn_17+ OPERTN_18+
opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24 like '%H20[16]%' or
                                opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+
OPERTN_06+opertn_07+ OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+
OPERTN_16+opertn_17+ OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24   like
'%H22%')
                                and ADMIDATE < '2012/01/04'
Order by Patient_ID, admidate
-----1196203
---creastin a temporary table with all endoscopies performed regardless of whether people have had a
diag of upper GI cancer and information associated with them
SET CONCAT_NULL_YIELDS_NULL OFF
select *
into #colonoscopies
from (select Patient_ID, SPELL_ID, SPELL_FINYR, EPIORDER,EPISTART, ADMIDATE, DISDATE, ADMIMETH,
CLASSPAT, DIAG_01, DIAG_02, DIAG_03, DIAG_04,DIAG_05, DIAG_06,
                DIAG_07, DIAG_08,DIAG_09, DIAG_10,DIAG_11, DIAG_12, DIAG_13, DIAG_14, DISMETH, GPPRAC,
MAINSPEF, OPERTN_01, OPERTN_02, OPERTN_03, OPERTN_04, OPERTN_05,
                OPERTN_06, OPERTN_07, opertn_08, OPERTN_09, OPERTN_10, OPERTN_11, OPERTN_12,
OPERTN_13, OPERTN_14, LEFT(procode,3) procode, SEX, STARTAGE, quintiles , resgor,
                case when opertn_01 like '%H20[16]%' OR OPERTN_01 like '%H22%' then OPDATE_01
                        when opertn_02 like '%H20[16]%' OR OPERTN_02 like '%H22%' then OPDATE_02
                        when opertn_03 like '%H20[16]%' OR OPERTN_03 like '%H22%' then OPDATE_03
                        when opertn_04 like '%H20[16]%' OR OPERTN_04 like '%H22%' then OPDATE_04
                        when opertn_05 like '%H20[16]%' OR OPERTN_05 like '%H22%' then OPDATE_05
                        when opertn_06 like '%H20[16]%' OR OPERTN_06 like '%H22%' then OPDATE_06
                        when opertn_07 like '%H20[16]%' OR OPERTN_07 like '%H22%' then OPDATE_07
                        when opertn_08 like '%H20[16]%' OR OPERTN_08 like '%H22%' then OPDATE_08
                        when opertn_09 like '%H20[16]%' OR OPERTN_09 like '%H22%' then OPDATE_09
                        when opertn_10 like '%H20[16]%' OR OPERTN_10 like '%H22%' then OPDATE_10
                        when opertn_11 like '%H20[16]%' OR OPERTN_11 like '%H22%' then OPDATE_11
                        when opertn_12 like '%H20[16]%' OR OPERTN_12 like '%H22%' then OPDATE_12
                        when opertn_13 like '%H20[16]%' OR OPERTN_13 like '%H22%' then OPDATE_13
                        when opertn_14 like '%H20[16]%' OR OPERTN_14 like '%H22%' then OPDATE_14
                        when opertn_15 like '%H20[16]%' OR OPERTN_15 like '%H22%' then OPDATE_15
                        when opertn_16 like '%H20[16]%' OR OPERTN_16 like '%H22%' then OPDATE_16
                        when opertn_17 like '%H20[16]%' OR OPERTN_17 like '%H22%' then OPDATE_17
                        when opertn_18 like '%H20[16]%' OR OPERTN_18 like '%H22%' then OPDATE_18
                        when opertn_19 like '%H20[16]%' OR OPERTN_19 like '%H22%' then OPDATE_19
```

```
                        when opertn_20 like '%H20[16]%' OR OPERTN_20 like '%H22%' then OPDATE_20 END
as colonoscopy_date

                from HED.dbo.HES_EPISODES_part1
                where (opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+ OPERTN_06+opertn_07+
OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+ OPERTN_16+opertn_17+
OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24 like '%H20[16]%' or
                                                opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+
OPERTN_06+opertn_07+ OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+
OPERTN_16+opertn_17+ OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24   like
'%H22%')
                                                and ADMIDATE > '2003/31/03'
                union ALL
                select Patient_ID, SPELL_ID, SPELL_FINYR, EPIORDER,EPISTART,  ADMIDATE, DISDATE,
ADMIMETH, CLASSPAT, DIAG_01, DIAG_02, DIAG_03, DIAG_04,DIAG_05, DIAG_06,
                DIAG_07, DIAG_08,DIAG_09, DIAG_10,DIAG_11, DIAG_12, DIAG_13, DIAG_14, DISMETH, GPPRAC,
MAINSPEF, OPERTN_01, OPERTN_02, OPERTN_03, OPERTN_04, OPERTN_05,
                OPERTN_06, OPERTN_07, opertn_08, OPERTN_09, OPERTN_10, OPERTN_11, OPERTN_12,
OPERTN_13, OPERTN_14, LEFT(procode,3) procode, SEX, STARTAGE, quintiles , resgor
                , case when opertn_01 like '%H20[16]%' OR OPERTN_01 like '%H22%' then OPDATE_01
                        when opertn_02 like '%H20[16]%' OR OPERTN_02 like '%H22%' then OPDATE_02
                        when opertn_03 like '%H20[16]%' OR OPERTN_03 like '%H22%' then OPDATE_03
                        when opertn_04 like '%H20[16]%' OR OPERTN_04 like '%H22%' then OPDATE_04
                        when opertn_05 like '%H20[16]%' OR OPERTN_05 like '%H22%' then OPDATE_05
                        when opertn_06 like '%H20[16]%' OR OPERTN_06 like '%H22%' then OPDATE_06
                        when opertn_07 like '%H20[16]%' OR OPERTN_07 like '%H22%' then OPDATE_07
                        when opertn_08 like '%H20[16]%' OR OPERTN_08 like '%H22%' then OPDATE_08
                        when opertn_09 like '%H20[16]%' OR OPERTN_09 like '%H22%' then OPDATE_09
                        when opertn_10 like '%H20[16]%' OR OPERTN_10 like '%H22%' then OPDATE_10
                        when opertn_11 like '%H20[16]%' OR OPERTN_11 like '%H22%' then OPDATE_11
                        when opertn_12 like '%H20[16]%' OR OPERTN_12 like '%H22%' then OPDATE_12
                        when opertn_13 like '%H20[16]%' OR OPERTN_13 like '%H22%' then OPDATE_13
                        when opertn_14 like '%H20[16]%' OR OPERTN_14 like '%H22%' then OPDATE_14
                        when opertn_15 like '%H20[16]%' OR OPERTN_15 like '%H22%' then OPDATE_15
                        when opertn_16 like '%H20[16]%' OR OPERTN_16 like '%H22%' then OPDATE_16
                        when opertn_17 like '%H20[16]%' OR OPERTN_17 like '%H22%' then OPDATE_17
                        when opertn_18 like '%H20[16]%' OR OPERTN_18 like '%H22%' then OPDATE_18
                        when opertn_19 like '%H20[16]%' OR OPERTN_19 like '%H22%' then OPDATE_19
                        when opertn_20 like '%H20[16]%' OR OPERTN_20 like '%H22%' then OPDATE_20 END
as colonoscopy_date
                from HED.dbo.HES_EPISODES
                where (opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+ OPERTN_06+opertn_07+
OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+ OPERTN_16+opertn_17+
OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24 like '%H20[16]%' or
                                                opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+
OPERTN_06+opertn_07+ OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+
OPERTN_16+opertn_17+ OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24   like
'%H22%')
                                                and ADMIDATE < '2009/01/04') t1
-----(1468549 row(s) affected)

---now identifiying patients with a diagnosis of upper GI cancer
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  Patient_ID, MIN(epistart) minadmission
into #LowerGIDiagnosis
from (select Patient_ID, epistart
        from HED.dbo.HES_EPISODES
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12+diag_13+ diag_14+diag_15+ diag_16+diag_17+ diag_18+diag_19+ diag_20 like
'%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12+diag_13+ diag_14+diag_15+ diag_16+diag_17+ diag_18+diag_19+ diag_20 like
'%c20%'
        union ALl
        select Patient_ID, epistart
        from HED.dbo.HES_EPISODES_part1
        where  diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12+diag_13+ diag_14+diag_15+ diag_16+diag_17+ diag_18+diag_19+ diag_20 like
'%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12+diag_13+ diag_14+diag_15+ diag_16+diag_17+ diag_18+diag_19+ diag_20 like
'%c20%') t1
group by PAtient_ID
----439249
```

```
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  Extract_HESID, Max(DOD) dod
into #LowerGIDiagnosisfromDeaths
from (select Extract_HESID, DOD
        from HES_DW_ONS.dbo.HED_ONS_Mortality
        where
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON_NEONATA
L_11+CAUSE_OF_DEATH_NON_NEONATAL_12+CAUSE_OF_DEATH_NON_NEONATAL_13
        +CAUSE_OF_DEATH_NON_NEONATAL_14+CAUSE_OF_DEATH_NON_NEONATAL_15 like '%c1[89]%' or
        CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON
_NEONATAL_11+CAUSE_OF_DEATH_NON_NEONATAL_12+CAUSE_OF_DEATH_NON_NEONATAL_13
        +CAUSE_OF_DEATH_NON_NEONATAL_14+CAUSE_OF_DEATH_NON_NEONATAL_15 like '%c20%'
        ) t1
group by Extract_HESID
-----(181830 row(s) affected)
---checking to see whether there are any patients who died of lower GI CAncer who weren't admitted to
hosp with it
select *
from  #LowerGIDiagnosisfromDeaths t1  left join #LowerGIDiagnosis t2
                                                    on t1.EXTRACT_HESID = t2.Patient_ID collate
database_default
where t2.Patient_ID is null
---there are 38114 patients here who died without a hospital admission owhere lower gi cancer was
recorded
---lets look at OP data
SET CONCAT_NULL_YIELDS_NULL OFF
select distinct  PatientID, MIN(apptdate) minadmission
into #LowerGIDiagnosisOP
from (select PatientID, apptdate
        from HES_DW.dbo.HES_OP_200607
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'
        union All
        select PatientID, apptdate
        from HES_DW.dbo.HES_OP_200708
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'
        union All
        select PatientID, apptdate
        from HES_DW.dbo.HES_OP_200809
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'
        union All
        select PatientID, apptdate
        from HES_DW.dbo.HES_OP_200910
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'
        union All
        select Patient_ID, apptdate
        from HES_DW.dbo.HES_OP_201011
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'
        union All
        select Patient_ID, apptdate
        from HES_DW.dbo.HES_OP_201112
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'
        union All
        select Patient_ID, apptdate
        from HES_DW.dbo.HES_OP_201213
```

```
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'
        union ALl
        select Patient_ID, apptdate
        from HES_DW.dbo.HES_OP_201314
        where diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c1[89]%' or
                        diag_01 + diag_02+diag_03+ diag_04+diag_05+ diag_06+diag_07+ diag_08+diag_09+
diag_10+diag_11+diag_12 like '%c20%'                                                ) t1
group by PAtientID
----(25020 row(s) affected)

select Patient_ID, MIN(minadmission)minadmission
Into #LowerGI
from (select Patient_ID, minadmission
                from #LowerGIDiagnosis
                union all
                select EXTRACT_HESID collate SQL_Latin1_General_CP1_CI_AS, DOD
                from #LowerGIDiagnosisfromDeaths
                union all
                select PatientID collate SQL_Latin1_General_CP1_CI_AS, minadmission
                from #LowerGIDiagnosisOP) a
group by Patient_ID
----482957

--truncate table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses;
---drop table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses;

---creating a table which only has the endoscopy info for the patients who have had a diagnosisi of
upper gi cancer at some point
SET CONCAT_NULL_YIELDS_NULL OFF
select t1.*, t2.minadmission firstAdmissionDiagnosis
into Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
from #colonoscopies t1 inner join #LowerGI t2
                                on t1.Patient_ID= t2.PAtient_ID

---(161359 row(s) affected)
----looking at the data
select *
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
---having looked at this it has confirmed that there are patients with endoscopys after their cancer
admission dates, we need to remove these rows
----checking whether there are any endoscopies performed after the diagnosis of upper gi cancer as we
won't be interested in these
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where epistart > firstAdmissionDiagnosis
---77167 rows
---and now deleting them
delete
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where epistart > firstAdmissionDiagnosis
----(77167 row(s) affected)


select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where colonoscopy_date not between ADMIDATE and  disdate
---(206 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set colonoscopy_date  = disdate
where colonoscopy_date not between ADMIDATE and  disdate
---(206 row(s) affected)

---adding in some additional variables that we will need for these analses
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add dayspostopColonoscopyDiag int,
LastColonoscopybeforeDiag int,
MissedDiag int;
```

```
---calculating the length of time from the endoscopy until the diagnosis of upper GI cancer - as this
will be needed to know whether the
---endoscopy was successful or not

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set dayspostopColonoscopyDiag = DATEDIFF(dd, colonoscopy_date, firstAdmissionDiagnosis);
---(84912 row(s) affected)
---Lets check out the results of this
select dayspostopColonoscopyDiag, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
group by dayspostopColonoscopyDiag
order by dayspostopColonoscopyDiag

---there are some nulls and also some negative numbers
select SPELL_FINYR, ADMIDATE, colonoscopy_date, firstAdmissionDiagnosis, dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where dayspostopColonoscopyDiag is null or dayspostopColonoscopyDiag < 0
order by Spell_finYr
--the nulls look to be where the endoscopy date is null
---these look all to be for the admitting epsiode as the diagnosis has been put onto the admission
fate,lets look to see wherther
----any of these people will be
select SPELL_FINYR, ADMIDATE, DISDATE, colonoscopy_date, firstAdmissionDiagnosis,
dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where ADMIDATE >  firstAdmissionDiagnosis and( dayspostopColonoscopyDiag is null or
dayspostopColonoscopyDiag < 0)
order by Spell_finYr

---Should may update the diagnosis date to the discharge date? but first looking in some more detail
at the dates
select SPELL_FINYR, ADMIDATE, disdate, colonoscopy_date, firstAdmissionDiagnosis,
dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where dayspostopColonoscopyDiag is null or dayspostopColonoscopyDiag < 0
order by Spell_finYr

---lets leave this as it is now and replace the nulls - using the discharge date as the endoscopy date

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set dayspostopColonoscopyDiag = DATEDIFF(dd, admidate, firstAdmissionDiagnosis)
where colonoscopy_date is null
---87 rows
---there are some disdates here which are null and should be updated

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set DISDATE = t2.disdate
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_SPells t2
                                                    on t1.Patient_ID = t2.PATIENT_ID
                                                    and t1.SPELL_ID = t2.SPELL_ID


-----(84099 row(s) affected)
drop table #colonoscopies
drop table #LowerGI
drop table #LowerGIDiagnosis
drop table #LowerGIDiagnosisfromDeaths
drop table #LowerGIDiagnosisOP

---lets for back to the endoscopy date and look to see whether any of the negative dates occur within
3 months of the diagnosis  - this woiuld have been considered a success
select  SPELL_FINYR, ADMIDATE, disdate, colonoscopy_date, firstAdmissionDiagnosis,  case when
ABS(dayspostopColonoscopyDiag) > 90 then 1 else 0 end,dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where dayspostopColonoscopyDiag is null or dayspostopColonoscopyDiag < 0
order by case when ABS(dayspostopColonoscopyDiag) > 90 then 1 else 0 end
---6891 rows

 ---there are 5 rows where the colonoscopy date and the admission date are more than 90 days apart but
in the same spell
 --lets have a look at these patients see what information is available o them

select Patient_ID, SPELL_ID, EPIORDER, Epistart, ADMIDATE, DISDATE, colonoscopy_date,
firstAdmissionDiagnosis, dayspostopColonoscopyDiag
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
```

```
where dayspostopColonoscopyDiag < -90
Order by Patient_ID
---this has returned 5 rows - I think that some of these will have to be removed, as the endoscopy has
occurred months after the original admission date, but we won't know
---when the diagnosis occurred

--maybe creat a drop record field now
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add droprecord int;
---and drop the 6 records that have been found

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where dayspostopColonoscopyDiag < -90
----(5 row(s) affected) - we cannot be sure when during their admission they were diagnosed with Lower
GI cancer

---updating the LastEndoscopybeforediag
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastColonoscopybeforediag = 0 ; ---(84192 row(s) affected)
---this variable identifies whether the endoscopy belonging to this row was the lsat one before the
diagnosis of uppper GI cancer was made
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastColonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PAtient_ID,
MAX(colonoscopy_date)end_Date

                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                        where droprecord is null and (colonoscopy_date <=
firstAdmissionDiagnosis or

                                        (EPISTART = firstAdmissionDiagnosis and
dayspostopColonoscopyDiag < 0))

                                        group by Patient_ID ,firstAdmissionDiagnosis

                                        ) t2
                                        on t1.Patient_ID = t2.PAtient_ID
                                        and t1.colonoscopy_date = t2.end_Date
----(75236 row(s) affected)
---check that this is working
select EPISTART, ADMIDATE, DISDATE, CLASSPAT, colonoscopy_date, firstAdmissionDiagnosis,
dayspostopColonoscopyDiag , LastColonoscopybeforeDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and dayspostopColonoscopyDiag < 0 ---and EPISTART <> firstAdmissionDiagnosis
Order by dayspostopColonoscopyDiag
-------this does seem to be doing the right thing
----checking the data
select Patient_ID, sum(LastColonoscopybeforeDiag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(LastColonoscopybeforeDiag)  desc
----130 patients more than one last colonoscopy before the diagnosis
---Lets check these patients out
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where droprecord is null
                                        group by Patient_ID
                                        having sum(LastColonoscopybeforeDiag) >1)
        and LastColonoscopybeforeDiag = 1
Order by Patient_ID, ADMIDATE
----ssoe of these are because of the multiple epsiodes
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastColonoscopybeforeDiag = 0
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  (select Patient_Id,
spell_Id, MIN(epiorder) minepi
```

```
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                where Patient_ID in (select Patient_ID

                                                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                                                                                where
droprecord is null

                                                                                                                group by
Patient_ID

                                                                                                                having
sum(LastColonoscopybeforeDiag) >1)

                                                                        and LastColonoscopybeforeDiag = 1

                                                group by  Patient_Id, spell_Id) t2
                                                on t1.Patient_ID = t2.Patient_ID
                                                and t1.SPELL_ID = t2.SPELL_ID
                                                and t1.EPIORDER = t2.minepi
----(207 row(s) affected)
---Lets check this has worked
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where droprecord is null
                                        group by Patient_ID
                                        having sum(LastColonoscopybeforeDiag) >1)
        and LastColonoscopybeforeDiag = 1
Order by Patient_ID, ADMIDATE
---IT has! there are some people here with a zero though fot he last colonoscop - this shoudl not be
the case

select PATIent_ID, MAX(Colonoscopy_date),firstAdmissionDiagnosis, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select PATIent_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where droprecord is null
                                        group by PATIent_ID
                                        having  MAX( LastColonoscopybeforeDiag) = 0)
group by PATIent_ID,firstAdmissionDiagnosis
Order by Patient_ID, firstAdmissionDiagnosis
----148 rows
---there are a couple of patietns here who have only one record for their macimum date - lets look in
more detail than we just have
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select PATIent_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where droprecord is null
                                        group by PATIent_ID
                                        having  MAX( LastColonoscopybeforeDiag) = 0)
Order by Patient_ID, ADMIDATE
---these pretty much all thave the same episode number, but different spell numbers
select PATIent_ID
into   #missingLastDates
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by PATIent_ID
having  MAX( LastColonoscopybeforeDiag) = 0
---148 rows

-----Lets try and see whether we can identify one row per patient
select t1.PAtient_ID, admidate, MAX(spell_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select t1.Patient_ID,
MAX(Admidate) as Maxadmi
```

```
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join #missingLastDates t2


                on t1.Patient_ID = t2.Patient_ID

                                    group by t1.Patient_ID) t2
                                            on t1.Patient_ID = t2.Patient_ID
                                            and t1.ADMIDATE = t2.Maxadmi
group by t1.PAtient_ID, admidate
order by t1.PAtient_ID, admidate
---this has worked we are down to one admission per person
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastColonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select t1.PAtient_ID,
admidate, MAX(spell_ID) maxspell

                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select t1.Patient_ID,
MAX(Admidate) as Maxadmi


                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join #missingLastDates t2


                                                    on t1.Patient_ID = t2.Patient_ID


                                                                        group by
t1.Patient_ID) t2

                                                                                on
t1.Patient_ID = t2.Patient_ID

                                                                                and
t1.ADMIDATE = t2.Maxadmi

                                    group by t1.PAtient_ID, admidate) t2
                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                        and t1.ADMIDATE =
t2.ADMIDATE
                                                                        and t1.SPELL_ID =
t2.maxspell

---150 rows affected
---so we have duplication here again
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID = 'E1AD5D1A2EC9FDFC80476C6B1F375D33'
Order by Admidate
---This appears to be a duplicated row - so lets drop it

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID = 'E1AD5D1A2EC9FDFC80476C6B1F375D33'
----(2 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastColonoscopybeforeDiag = 0
where Patient_ID = 'D16FFCE9195537918588B446C5FDD22F' and SPELL_ID = 60762756647 and EPIORDER =11
----(1 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastColonoscopybeforeDiag = 0
where Patient_ID = 'EC44AF2C9B72104EA32CDB3B7882282C' and SPELL_ID = 708118474565 and EPIORDER =2
----(1 row(s) affected)

---lets look at the endoscopys which are more than five years before the diagnosis
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where dayspostopColonoscopyDiag >1826
-----6638 rows
```

```
---adding a nother variable to the data - diagnosis this variable will be a one where the endoscopy
occurred within 90 days of the diagnosis
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add diagnosis int;

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set diagnosis =1
where dayspostopColonoscopyDiag between 0 and  90
----(55333 row(s) affected)
---and set the diagnosis = 1 where the diagnosis ocurred in the same spell as the endoscopy
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and diagnosis is null and dayspostopColonoscopyDiag < 0
--6886
---and update these records
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set diagnosis =1
where droprecord is null and diagnosis is null and dayspostopColonoscopyDiag < 0
-----(6886 row(s) affected)

---lets update the missed diagnoses
Update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set MissedDiag = 1
where dayspostopColonoscopyDiag between 91 and 1826
 ----(15330 row(s) affected)

 ---check this is doing what it should be
 select *
 from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
 where MissedDiag = 1

 ----I need to update the discharge method to that of the spells

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set DISMETH = t2.dismeth
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_SPElls t2
                                                                        on t1.Patient_ID =
t2.PATIENT_ID
                                                                  and t1.SPELL_ID = t2.SPELL_ID
-----(84099 row(s) affected)

---lets look at cleaning the data - first checking the genders are correct

select sex, COUNT(patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
group by sex

---there are 5 people with an invalid sex
---lets look at these patietns
select Patient_ID, admidate
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where SEX in (0,9)
---5 rows
select *
from hed.dbo.hes_spells
where PATIENT_ID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2)
---no results - lets have a look to see whether any of these patients have been to outpatients and
have a valid gender

select *
from HES_DW.dbo.HES_OP_200607
where PATIENTID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2);
        ---no
select *
from HES_DW.dbo.HES_OP_200708
where PATIENTID in (select Patient_ID
```

```
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2); --no

select *
from HES_DW.dbo.HES_OP_200809
where PATIENTID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2);

select *
from HES_DW.dbo.HES_OP_200910
where PATIENTID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2);

select *
from HES_DW.dbo.HES_OP_201011
where PATIENT_ID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2);

select *
from HES_DW.dbo.HES_OP_201112
where PATIENT_ID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2);

---or deaths
select *
from HES_DW_ONS.dbo.HED_ONS_Mortality
where EXTRACT_HESID in (select Patient_ID collate SQL_Latin1_General_CP1_CI_AS
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where SEX in (0,9))
        and sex in (1,2);
        ----one patient
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set sex = 1
where Patient_ID =      '9AA94D7CCDD6062531AE3CC4EF74191F'

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord =1
where SEX in (0,9)
----(4 row(s) affected)

---moving on to checking the GOR codes are withing England
select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
group by resgor
order by resgor

---there are 12 patients here of no fixed  abode
---there are in total 370 patients with a GOR which is not valid, lets look at these patients and see
whether they have been admitted
--- another time in the for the same groups with a fixed address
select resgor, COUNT(distinct Patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where resgor in ('u','x','y','z')
                                        )
         and resgor in ('a','b','c','d','e','f','g','h','i','k','j')
group by resgor
```

```
order by resgor
---about 6 are...
select distinct Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where resgor in ('u','x','y','z')
                                        )
         and resgor in ('a','b','c','d','e','f','g','h','i','k','j')
---16 rows
----    lets have a look at these patients
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select distinct Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where Patient_ID in (select Patient_ID
                                                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                                                where resgor
in ('u','y','z')
                                                                                                                )
                                                                                                and resgor in
('a','b','c','d','e','f','g','h','i','k','j'))
order by Patient_ID, admidate

---lets leave the 'u's as they are and update the y's and z's

----I think we'll have to drop the records where a patient has been listed as living in Scotland,
Wales or abroad as they may have had other admissions in these countries
---that are not recorded in HES
select distinct Patient_ID
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where resgor in ('s','w','x')
---238 rows

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in (select distinct Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where resgor in ('s','w','x'))
----(267 row(s) affected)

select distinct t1.PAtient_ID, t1.resgor, t2.resgor
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
Patient_ID, resgor

                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                        where resgor in
('a','b','c','d','e','f','g','h','i','k','j')) t2
                                                                                                on
t1.patient_ID = t2.patient_ID
where t1.resgor in ('y','z')
order by PAtient_ID
---14 rows - check these are disinct patients

select distinct patient_ID
from (select distinct t1.PAtient_ID, t1.resgor, t2.resgor resgornew
                        from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
(select distinct Patient_ID, resgor

                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                                        where resgor in
('a','b','c','d','e','f','g','h','i','k','j')) t2

                on t1.patient_ID = t2.patient_ID
                        where t1.resgor in ('y','z'))  a
---14 so this is correct
```

```
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor  = t2.resgornew
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.PAtient_ID, t1.resgor, t2.resgor resgornew

                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct Patient_ID,
resgor


                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


                                                                                    where resgor
in ('a','b','c','d','e','f','g','h','i','k','j')) t2


                                          on t1.patient_ID = t2.patient_ID

                                                  where t1.resgor in ('y','z')) t2
                                                                        on t1.Patient_ID=
t2.PAtient_ID
                                                                        and t1.resgor =
t2.resgor
-----(16 row(s) affected)

---lets relook at the data now
select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord  is null
group by resgor
order by resgor
---there are still about 194 patients with an invalid GOR code
---lets check these patients back to the HES tables
select t1.Patient_ID, t1.spell_ID, t1.resgor, t2.resgor
from HED.dbo.HES_Episodes t1 inner join (select resgor, Patient_ID, spell_ID
                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                        where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
                                                        and t1.SPELL_ID = t2.SPELL_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---this has resulted in 6  further patients with valid resgor

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor  = t2.resgornew
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1      inner join (select
t1.Patient_ID, t1.spell_ID, t1.resgor resgornew, t2.resgor

                                                        from HED.dbo.HES_Episodes t1 inner join
(select resgor, Patient_ID, spell_ID


                                  from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


                                          where droprecord  is null and resgor in ('y','z')) t2


        on t1.PATIENT_ID = t2.Patient_ID


                and t1.SPELL_ID = t2.SPELL_ID

                                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                        and t1.SPELL_ID =
t2.SPELL_ID
----(4 row(s) affected)
```

```
select t1.Patient_ID, t1.spell_ID, t1.resgor, t2.resgor
from HED.dbo.HES_Episodes_part1 t1 inner join (select resgor, Patient_ID, spell_ID
                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
                                                        and t1.SPELL_ID = t2.SPELL_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---nothing

        ---lets look to see whether there are any relevant GOR codes within 90 days of the admission

select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes_part1 t1 inner join (select resgor, Patient_ID, Admidate
                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
                                                        and abs(DATEDIFF(dd, t1.Admidate,
t2.Admidate)) between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---10 rows
---and now update these patients
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.Patient_ID, t1.resgor resgornew, t2.resgor

                                                from HED.dbo.HES_Episodes_part1 t1 inner join (select
resgor, Patient_ID, Admidate


                                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                                on
t1.PATIENT_ID = t2.Patient_ID


        and abs(DATEDIFF(dd, t1.Admidate, t2.Admidate)) between 0 and 90

                                                where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                on t1.Patient_ID =
t2.Patient_ID
                                                                                and t1.resgor =
t2.resgor
---11 row(s) affected)
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes t1 inner join (select resgor, Patient_ID, Admidate
                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
                                                        and abs(DATEDIFF(dd, t1.Admidate,
t2.Admidate)) between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---44 rows
---and update
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.Patient_ID, t1.resgor resgornew, t2.resgor

                                                from HED.dbo.HES_Episodes t1 inner join (select
resgor, Patient_ID, Admidate


                                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
```

```
                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                                    on
t1.PATIENT_ID = t2.Patient_ID


        and abs(DATEDIFF(dd, t1.Admidate, t2.Admidate)) between 0 and 90

                                                    where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                                        and t1.resgor =
t2.resgor
----(44 row(s) affected)
--and now check how this has worked
select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord  is null
group by resgor
order by resgor
---down to 72 - lets look at outpatients for these patietns
select distinct t1.PatientID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_200708 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENTID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---17 rowsa here which we can update


update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.PatientID, t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_200708 t1 inner join (select resgor,
Patient_ID, Admidate


                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                where droprecord  is null and resgor in ('y','z')) t2

                                                                                        on t1.PATIENTID =
t2.Patient_ID

                                                                                                    and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                        on t1.Patient_ID =
t2.PatientID
                                                                                        and t1.resgor =
t2.resgor;
----(17 row(s) affected)

select distinct t1.PatientID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_200809 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENTID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
```

```
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')

--1 row - lets update this manually
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor = 'K'
where Patient_ID = 'A1FD6B09B87E68B033FC43B3536DF51E' and resgor = 'Y'

---(1 row(s) affected)
select distinct t1.PatientID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_200910 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENTID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---update the table
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor = 'H'
where Patient_ID in  ('34D3E70B48112B582A3ACE4B636946AD') and resgor = 'Y'
---(1 row(s) affected)

select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_201011 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENT_ID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
--- nothing
---last one
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_201112 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENT_ID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
----nothing
----lets check this again
select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord  is null
group by resgor
order by resgor
---53 rows whcih are still y or z
----lets just look to see whether there are any records

select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes t1 inner join (select resgor, Patient_ID, Admidate
                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---23 rows here
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes_Part1 t1 inner join (select resgor, Patient_ID, Admidate
                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---27 rows maybe just update these
```

```
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.Patient_ID, t1.resgor resgornew, t2.resgor

                                        from HED.dbo.HES_Episodes t1 inner join (select
resgor, Patient_ID, Admidate

                                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                        on
t1.PATIENT_ID = t2.Patient_ID

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                        and t1.resgor =
t2.resgor
---(23 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.Patient_ID, t1.resgor resgornew, t2.resgor

                                        from HED.dbo.HES_Episodes_part1 t1 inner join (select
resgor, Patient_ID, Admidate

                                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                        on
t1.PATIENT_ID = t2.Patient_ID



                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                        and t1.resgor =
t2.resgor
----(11 row(s) affected)

---the ones that are left can't be updated so may need to be removed

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where resgor in ('y','z')
----(20 row(s) affected)
---check the spell year

select SPELL_FINYR,droprecord, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
group by SPELL_FINYR,droprecord
order by SPELL_FINYR     ,droprecord
---there are 14 patients where the spell year is 1000  and they currently have not been dropped
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where SPELL_FINYR   = '1000'
Order by Patient_ID, ADMIDATE
---93 rows
---lets link these back to the hes tables

select t2.*
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
spell_ID, epiorder, admidate, disdate, epistart, epiend, epistat, SPELL_FINYR
```

```
                                        from HED.dbo.HES_EPISODES

                                    union all

                                        select Patient_ID, spell_ID, epiorder, admidate, disdate,
epistart, epiend, epistat, SPELL_FINYR

                                        from HED.dbo.HES_EPISODES_Part1) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.SPELL_ID = t2.SPELL_ID
                                                                                    and
t1.EPIORDER = t2.Epiorder
where t1.SPELL_FINYR   = '1000'
Order by Patient_ID, ADMIDATE
---93 rows
---all these episodes have finished so I'm not sure what has gone on - it is something to do with the
date of discharge, lets look to see whether there are any epsiodes in the same spell with a
---date of discharge
select t2.*
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
spell_ID, epi_last, epiorder, admidate, disdate, epistart, epiend, epistat, DISMETH, SPELL_FINYR

                                        from HED.dbo.HES_EPISODES

                                    union all

                                        select Patient_ID, spell_ID, epiorder, epi_last,  admidate,
disdate, epistart, epiend, epistat, DISMETH,SPELL_FINYR

                                        from HED.dbo.HES_EPISODES_Part1) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.SPELL_ID = t2.SPELL_ID

where t1.SPELL_FINYR   = '1000'
Order by dismeth
---probably best to remove these records, for the most part it looks as if these spells have not
finished
--206
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where SPELL_FINYR = '1000'
---(93 row(s) affected)

-----Look at admimeth
select admimeth, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord  is null
group by admimeth
order by admimeth

---look to see whether there is any difference between this and the spells table
select t1.Patient_ID, t1.Admimeth, t2.Admimeth
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_SPELLS t2
                                                                                    on
t1.Patient_ID = t2.PATIENT_ID
                                                                                    and
t1.SPELL_ID = t2.Spell_ID
where t1.ADMIMETH > 90 and t2.ADMIMETH <99
---nope
---Now using the episodes table
select t1.Patient_ID, t1.Admimeth, t2.Admimeth
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_Episodes_part1 t2
                                                                                    on
t1.Patient_ID = t2.PATIENT_ID
                                                                                    and
t1.SPELL_ID = t2.Spell_ID
where t1.ADMIMETH > 90 and t2.ADMIMETH <99
---again, no
```

```
----not sure what to do with these records -  I don't suppose it will matter too much
---check disdate
select disdate, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord  is null
group by disdate
order by disdate
---this looks OK no nulls
---move onto dismeth
select dismeth, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord  is null
group by dismeth
order by dismeth
---there are 22 here with validation errors - I'm not sure what to do with these either!
---lets look at the class pats
select classpat, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord  is null
group by classpat
order by classpat
---we've got some regular day attenders here ....
---lets loook at these patients
select *
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and CLASSPAT =3
order by Patient_ID

---43 rows

----Lets have a look at the number of people admitted by the year of their first admission
select COUNT(distinct Patient_ID) , case when firstAdmissionDiagnosis between convert(datetime,
'2001/01/04') and CONVERT(datetime, '2002/31/03') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/01/04') and CONVERT(datetime, '2003/31/03')
then '2002/03'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/01/04') and CONVERT(datetime, '2004/31/03')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/01/04') and CONVERT(datetime, '2005/31/03')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/01/04') and CONVERT(datetime, '2006/31/03')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/01/04') and CONVERT(datetime, '2007/31/03')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/01/04') and CONVERT(datetime, '2008/31/03')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/01/04') and CONVERT(datetime, '2009/31/03')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/01/04') and CONVERT(datetime, '2010/31/03')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/01/04') and CONVERT(datetime, '2011/31/03')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/01/04') and CONVERT(datetime, '2012/31/03')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/01/04') and CONVERT(datetime, '2013/31/03')
then '2012/13' end fin_yr
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by case when firstAdmissionDiagnosis between convert(datetime, '2001/01/04') and
CONVERT(datetime, '2002/31/03') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/01/04') and CONVERT(datetime, '2003/31/03')
then '2002/03'
```

```
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/01/04') and CONVERT(datetime, '2004/31/03')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/01/04') and CONVERT(datetime, '2005/31/03')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/01/04') and CONVERT(datetime, '2006/31/03')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/01/04') and CONVERT(datetime, '2007/31/03')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/01/04') and CONVERT(datetime, '2008/31/03')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/01/04') and CONVERT(datetime, '2009/31/03')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/01/04') and CONVERT(datetime, '2010/31/03')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/01/04') and CONVERT(datetime, '2011/31/03')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/01/04') and CONVERT(datetime, '2012/31/03')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/01/04') and CONVERT(datetime, '2013/31/03')
then '2012/13' end
order by case when firstAdmissionDiagnosis between convert(datetime, '2001/01/04') and
CONVERT(datetime, '2002/31/03') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/01/04') and CONVERT(datetime, '2003/31/03')
then '2002/03'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/01/04') and CONVERT(datetime, '2004/31/03')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/01/04') and CONVERT(datetime, '2005/31/03')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/01/04') and CONVERT(datetime, '2006/31/03')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/01/04') and CONVERT(datetime, '2007/31/03')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/01/04') and CONVERT(datetime, '2008/31/03')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/01/04') and CONVERT(datetime, '2009/31/03')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/01/04') and CONVERT(datetime, '2010/31/03')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/01/04') and CONVERT(datetime, '2011/31/03')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/01/04') and CONVERT(datetime, '2012/31/03')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/01/04') and CONVERT(datetime, '2013/31/03')
then '2012/13' end
---there are also a few nulls here - so we probably need to check this out as wll 1724 of them!
---Lets have a look at the nulls from earlier
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and  case when firstAdmissionDiagnosis between convert(datetime,
'2001/01/04') and CONVERT(datetime, '2002/31/03') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/01/04') and CONVERT(datetime, '2003/31/03')
then '2002/03'
```

```
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/01/04') and CONVERT(datetime, '2004/31/03')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/01/04') and CONVERT(datetime, '2005/31/03')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/01/04') and CONVERT(datetime, '2006/31/03')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/01/04') and CONVERT(datetime, '2007/31/03')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/01/04') and CONVERT(datetime, '2008/31/03')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/01/04') and CONVERT(datetime, '2009/31/03')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/01/04') and CONVERT(datetime, '2010/31/03')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/01/04') and CONVERT(datetime, '2011/31/03')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/01/04') and CONVERT(datetime, '2012/31/03')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/01/04') and CONVERT(datetime, '2013/31/03')
then '2012/13' end  is null
order by Patient_ID, admidate
---these are for patietns whose first admission happends in 2013
---lets have a look to see whether we are dropping records in one year more than in any other
Select case when firstAdmissionDiagnosis between convert(datetime, '2001/01/04') and CONVERT(datetime,
'2002/31/03') then '2001/02'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/01/04') and CONVERT(datetime, '2003/31/03')
then '2002/03'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/01/04') and CONVERT(datetime, '2004/31/03')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/01/04') and CONVERT(datetime, '2005/31/03')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/01/04') and CONVERT(datetime, '2006/31/03')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/01/04') and CONVERT(datetime, '2007/31/03')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/01/04') and CONVERT(datetime, '2008/31/03')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/01/04') and CONVERT(datetime, '2009/31/03')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/01/04') and CONVERT(datetime, '2010/31/03')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/01/04') and CONVERT(datetime, '2011/31/03')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/01/04') and CONVERT(datetime, '2012/31/03')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/01/04') and CONVERT(datetime, '2013/31/03')
then '2012/13' end fin_yr,
                                                                    sum(droprecord)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
group by case when firstAdmissionDiagnosis between convert(datetime, '2001/01/04') and
CONVERT(datetime, '2002/31/03') then '2001/02'
```

```
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2002/01/04') and CONVERT(datetime, '2003/31/03')
then '2002/03'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2003/01/04') and CONVERT(datetime, '2004/31/03')
then '2003/04'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2004/01/04') and CONVERT(datetime, '2005/31/03')
then '2004/05'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2005/01/04') and CONVERT(datetime, '2006/31/03')
then '2005/06'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2006/01/04') and CONVERT(datetime, '2007/31/03')
then '2006/07'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2007/01/04') and CONVERT(datetime, '2008/31/03')
then '2007/08'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2008/01/04') and CONVERT(datetime, '2009/31/03')
then '2008/09'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2009/01/04') and CONVERT(datetime, '2010/31/03')
then '2009/10'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2010/01/04') and CONVERT(datetime, '2011/31/03')
then '2010/11'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2011/01/04') and CONVERT(datetime, '2012/31/03')
then '2011/12'
                                                                    when
firstAdmissionDiagnosis between convert(datetime, '2012/01/04') and CONVERT(datetime, '2013/31/03')
then '2012/13' end
                                                                    ---nope, this seems
to be consistent

select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null

----Lets add dod and causes of death to this table
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add dod datetime,
underlying varchar(5),
cause_1 varchar(5),
cause_2 varchar(5),
cause_3 varchar(5);

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set dod = t2.DOD
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Extract_HeSID,
MAX(DOD) DOD

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality

                                        group by EXTRACT_HESID) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
----(46155 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set underlying = t2.CAUSE_OF_DEATH
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,

                                                t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD


                                    from HES_DW_ONS.dbo.HED_ONS_Mortality
```

```
                                group by EXTRACT_HESID) t2

                                        on t1.EXTRACT_HESID = t2.EXTRACT_HESID

                                        and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
----(46155 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set cause_1 = t2.CAUSE_OF_DEATH_NON_NEONATAL_1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,

                                                t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD


                                from HES_DW_ONS.dbo.HED_ONS_Mortality


                                group by EXTRACT_HESID) t2

                                        on t1.EXTRACT_HESID = t2.EXTRACT_HESID

                                        and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
---(46155 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set cause_2 = t2.CAUSE_OF_DEATH_NON_NEONATAL_2
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,

                                                t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD


                                from HES_DW_ONS.dbo.HED_ONS_Mortality


                                group by EXTRACT_HESID) t2

                                        on t1.EXTRACT_HESID = t2.EXTRACT_HESID

                                        and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
----    (46155 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set cause_3 = t2.CAUSE_OF_DEATH_NON_NEONATAL_3
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
t1.EXTRACT_HESID, t1.DOD, t1.CAUSE_OF_DEATH, t1.CAUSE_OF_DEATH_NON_NEONATAL_1,

                                                t1.CAUSE_OF_DEATH_NON_NEONATAL_2,
t1.CAUSE_OF_DEATH_NON_NEONATAL_3

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join (select
Extract_HeSID, MAX(DOD) DOD


                                from HES_DW_ONS.dbo.HED_ONS_Mortality


                                group by EXTRACT_HESID) t2

                                        on t1.EXTRACT_HESID = t2.EXTRACT_HESID
```

```
                                        and t1.DOD = t2.DOD) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
-----   (46155 row(s) affected)

----lets have a look at the data I've just added in
select Patient_ID, SPELL_FINYR, ADMIDATE, DISDATE,dismeth,firstAdmissionDiagnosis,  DOD,underlying,
cause_1, cause_2, cause_3
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where DOD is not null


---check whether there are any deaths recorded as happening before the first record of GI cancer
select Patient_ID, SPELL_FINYR, ADMIDATE, DISDATE,dismeth,firstAdmissionDiagnosis,  DOD,underlying,
cause_1, cause_2, cause_3
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where DOD is not null    and DOD < firstAdmissionDiagnosis
--- 11 rows, and most of these the DOD is a  long time before the cancer aignosis
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where  DOD is not null   and DOD < firstAdmissionDiagnosis
-----(11 row(s) affected)

--lets just check the ages of these paitnets
select startAge, COUNT(distinct Patient_ID)
from (select Patient_ID, MIN(startAge) StartAgE
                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                group by Patient_ID) t1
group by StartAgE
order by StartAgE

---10 patietns don't seem to have a startage, there are also 20 patients under 13
---Lets look at the patients who do not have a start age
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where STARTAGE is null
---10 rows
---quite a lot of these have a drop record
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where STARTAGE is null  and droprecord is null
---only 6 rows here
select Patient_ID, StartAge
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Patient_ID in (select distinct Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where STARTAGE is null  and droprecord is null )
---no additional age groups here...
---Lets have a look to see whether there are any ages for these patietns on HES

select t1.Patient_ID, t1.StartAge, ADMIAGE, ACTIVAGE, ENDAGE
from HED.dbo.HES_Episodes t1 inner join (select distinct Patient_ID, spell_ID
                                                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                                        where
STARTAGE is null  and droprecord is null )t2
                                        on t1.Patient_ID = t2.Patient_ID
                                        and t1.SPELL_ID = t2.SPELL_ID
----lets update these now...
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set STARTAGE = Admiage
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select t1.Patient_ID,
t1.StartAge, ADMIAGE, ACTIVAGE, ENDAGE

                                        from HED.dbo.HES_Episodes t1 inner join (select distinct
Patient_ID, spell_ID


                                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


                                where STARTAGE is null  and droprecord is null )t2
```

```
                                                                    on t1.Patient_ID =
t2.Patient_ID

                                                                    and t1.SPELL_ID =
t2.SPELL_ID) t2
                                                                    on t1.Patient_ID =
t2.Patient_ID
where t1.STARTAGE is null
----(5 row(s) affected)

---have a look in the first hes episodes tab;e
select t1.Patient_ID, t1.StartAge, ADMIAGE, ACTIVAGE, ENDAGE
from HED.dbo.HES_Episodes_Part1 t1 inner join (select distinct Patient_ID, spell_ID
                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                    where
STARTAGE is null  and droprecord is null )t2
                                        on t1.Patient_ID = t2.Patient_ID
                                        and t1.SPELL_ID = t2.SPELL_ID ---nope nothing
---lets remove the rest of the records with a null start age

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where STARTAGE is null or STARTAGE = 999
----(7 row(s) affected)
---after speaking to Ravi he suggested that the best hting to do would be to remove those patients
under 18
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in (select distinct Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                        where STARTAGE < 18 or STARTAGE > 200)---35 rows
-----(37 row(s) affected)



---Need to add some additional factors here looking at chemo/radiotherapy/surgery and ethnicity
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add chemo int,
radiotherapy int,
Surgery varchar(50),
ethnos varchar(1),
Renal_disease int,
Acute_MI int,
Cerebral_vasc_acc int,
CHF int,
Conn_tissue_disorder int,
Dementia int,
Diabetes int,
Liver_disease int,
Peptic_Ulcer int,
Peripheral_vasc_disease int,
Pulmonary_disease int,
Cancer int,
Diabetes_complications int,
Paraplegia int,
Metastatic_cancer int,
HIV int,
Severe_liver_dis int;
---lets update the records to indicate wehtehr or not people have had chemo - this would probably be
better done inan individual table
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 0 ;

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                        from HED.dbo.HES_episodes
```

```
                                    where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                            or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                            or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                            or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.epistart
----(15273 row(s) affected)
----- and the first part of the HES tabl

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                    from HED.dbo.HES_episodes_part1

                                    where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                            or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                            or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                            or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.epistart
---(8184 row(s) affected)
----lets  look to see if we can get any more infor off of the ICD 10 codes
SET CONCAT_NULL_YIELDS_NULL OFF
select t1.*
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                    from HED.dbo.HES_episodes_part1

                                    where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z082%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z511%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z542%') t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.epistart
```

```
Order by chemo
---there are some here where there isn't a record of chemo

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                        from HED.dbo.HES_episodes_part1

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z082%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z511%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z542%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.epistart
----(9200 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                        from HED.dbo.HES_episodes

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z082%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z511%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z542%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.epistart
---(15460 row(s) affected

---lets have a look to see wht the most commmonly recorded surgeries are
select  LEFT(t2.opertn_01, 3) , COUNT(distinct t2.Patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_Episodes t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.admidate
group by        LEFT(t2.opertn_01, 3)
Order by  COUNT(distinct t2.Patient_ID)
---1081 rows - H22, H33, H20, H07, X70, H25,  etc...
---think hte best way to deal with this is to get NT to confirm the procedures that they wnt to look
at

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Surgery  = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Epistart

                                        from HED.dbo.HES_episodes_part1

                                        where
```

```
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[45][12389]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[67][123489]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[89][1234589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H1[01][1234589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H12[2389]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H29[123489]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H33[123456789]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H34[1589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H40[123]%'


                                        ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Epistart
---(27841 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Surgery  = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                        from HED.dbo.HES_episodes

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[45][12389]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[67][123489]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[89][1234589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H1[01][1234589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H12[2389]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H29[123489]%'
```

```
                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H33[123456789]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H34[1589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H40[123]%'

                                        ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.epistart
---(30599 row(s) affected)

alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add resection tinyint;

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resection  = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Epistart

                                        from HED.dbo.HES_episodes_part1

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[45][12389]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[67][123489]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[89][1234589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H1[01][1234589]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H29[123489]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H33[123456789]%'

                                                    or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H34[189]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H122%'


                                        ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Epistart
---(27792 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resection  = 1
```

```
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                        from HED.dbo.HES_episodes

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[45][12389]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[67][123489]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H0[89][1234589]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H1[01][1234589]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H122%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H29[123489]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H33[123456789]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H34[189]%'



                                        ) t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.epistart
-----(30527 row(s) affected)
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add resection_date datetime;
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resection_date   = case when t3.OP_DATE is not null then t3.OP_DATE else t2.admidate end
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_SPELLS t2

        on t1.Patient_ID = t2.Patient_ID collate database_default

           inner join HED.dbo.HES_OPER_EXTRACT t3

                                on t2.spell_ID = t3.spell_ID
where t3.OPER_CODE like '%H0[45][12389]%' or  t3.OPER_CODE like '%H0[67][123489]%' or  t3.OPER_CODE
like  '%H0[89][1234589]%' or  t3.OPER_CODE like '%H1[01][1234589]%'
                                 or  t3.OPER_CODE like    '%H122%' or  t3.OPER_CODE like
'%H29[123489]%' or  t3.OPER_CODE like '%H33[123456789]%' or  t3.OPER_CODE like     '%H34[189]%'
------(31072 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set resection_date   = case when t3.OP_DATE is not null then t3.OP_DATE else t2.admidate end
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_SPELLS t2

        on t1.Patient_ID = t2.Patient_ID collate database_default

           inner join HED.dbo.HES_OPER_EXTRACT_part1 t3

                                on t2.spell_ID = t3.spell_ID
where t3.OPER_CODE like '%H0[45][12389]%' or  t3.OPER_CODE like '%H0[67][123489]%' or  t3.OPER_CODE
like  '%H0[89][1234589]%' or  t3.OPER_CODE like '%H1[01][1234589]%'
```

```
                                or  t3.OPER_CODE like    '%H122%' or  t3.OPER_CODE like
'%H29[123489]%' or  t3.OPER_CODE like '%H33[123456789]%' or  t3.OPER_CODE like     '%H34[189]%'
-----(28828 row(s) affected)

----updating the comorbid conditions
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Acute_MI = case when diag_01 like 'I2[123]%' or diag_02 like 'I2[123]%' or diag_03 like 'I2[123]%'
or diag_04 like 'I2[123]%'  or diag_05 like 'I2[123]%'
                 or diag_06 like 'I2[123]%'  or diag_07 like 'I2[123]%'  or diag_08 like 'I2[123]%'
or diag_09 like 'I2[123]%'  or diag_10 like 'I2[123]%'
                  or diag_11 like 'I2[123]%'  or diag_12 like 'I2[123]%'  or diag_13 like 'I2[123]%'
or diag_14 like 'I2[123]%'   then 1
        WHEN  diag_01 like 'I25[28]%' or diag_02 like 'I25[28]%' or diag_03 like 'I25[28]%' or diag_04
like 'I25[28]%' or diag_05 like 'I25[28]%' or diag_06 like 'I25[28]%' or
        diag_07 like 'I25[28]%' or diag_08 like 'I25[28]%' or diag_09 like 'I25[28]%' or diag_10 like
'I25[28]%' or diag_11 like 'I25[28]%' or diag_12 like 'I25[28]%' or
        diag_13 like 'I25[28]%' or diag_14 like 'I25[28]%'        then 1 else 0 end ;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Cerebral_vasc_acc = case when diag_01 like 'I6%' or diag_02 like 'I6%' or diag_03 like 'I6%'  or
diag_04 like 'I6%'  or diag_05 like 'I6%'
                 or diag_06 like 'I6%'  or diag_07 like 'I6%'  or diag_08 like 'I6%'  or diag_09 like
'I6%'  or diag_10 like 'I6%'
                  or diag_11 like 'I6%'  or diag_12 like 'I6%'  or diag_13 like 'I6%'  or diag_14 like
'I6%'  then 1
        WHEN  diag_01 like 'G45[012489]%' or diag_02 like 'G45[012489]%' or diag_03 like
'G45[012489]%' or diag_04 like 'G45[012489]%' or diag_05 like 'G45[012489]%' or diag_06 like
'G45[012489]%' or
                diag_07 like 'G45[012489]%' or diag_08 like 'G45[012489]%' or diag_09 like
'G45[012489]%' or diag_10 like 'G45[012489]%' or diag_11 like 'G45[012489]%' or diag_12 like
'G45[012489]%' or
                diag_13 like 'G45[012489]%' or diag_14 like 'G45[012489]%' then 1
        WHEN  diag_01 like 'G46%' or diag_02 like 'G46%' or diag_03 like 'G46%' or diag_04 like 'G46%'
or diag_05 like 'G46%' or diag_06 like 'G46%' or
                diag_07 like 'G46%' or diag_08 like 'G46%' or diag_09 like 'G46%' or diag_10 like
'G46' or diag_11 like 'G46%' or diag_12 like 'G46%' or
                diag_13 like 'G46%' or diag_14 like 'G46%'        then 1 else 0 end;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set CHF = case when diag_01 like 'I50%' or diag_02 like 'I50%' or diag_03 like 'I50%'  or diag_04 like
'I50%'  or diag_05 like 'I50%'
                 or diag_06 like 'I50%'  or diag_07 like 'I50%'  or diag_08 like 'I50%'  or diag_09
like 'I50%'  or diag_10 like 'I50%'
                  or diag_11 like 'I50%'  or diag_12 like 'I50%'  or diag_13 like 'I50%'  or diag_14
like 'I50%'  then 1
         else 0 end ;
        -----(84192 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Conn_tissue_disorder = case when diag_01 like 'M05%' or diag_02 like 'M05%' or diag_03 like 'M05%'
or diag_04 like 'M05%'  or diag_05 like 'M05%'
                 or diag_06 like 'M05%'  or diag_07 like 'M05%'  or diag_08 like 'M05%'  or diag_09
like 'M05%'  or diag_10 like 'M05%'
                  or diag_11 like 'M05%'  or diag_12 like 'M05%'  or diag_13 like 'M05%'  or diag_14
like 'M05%'   then 1
        WHEN  diag_01 like 'M06[039]%' or diag_02 like 'M06[039]%' or diag_03 like 'M06[039]%' or
diag_04 like 'M06[039]%' or diag_05 like 'M06[039]%' or diag_06 like 'M06[039]%' or
                diag_07 like 'M06[039]%' or diag_08 like 'M06[039]%' or diag_09 like 'M06[039]%' or
diag_10 like 'M06[039]%' or diag_11 like 'M06[039]%' or diag_12 like 'M06[039]%' or
                diag_13 like 'M06[039]%' or diag_14 like 'M06[039]%'  then 1
        WHEN  diag_01 like 'M3[24]' or diag_02 like 'M3[24]%' or diag_03 like 'M3[24]%' or diag_04
like 'M3[24]%' or diag_05 like 'M3[24]%' or diag_06 like 'M3[24]%' or
                diag_07 like 'M3[24]%' or diag_08 like 'M3[24]%' or diag_09 like 'M3[24]%' or diag_10
like 'M3[24]' or diag_11 like 'M3[24]%' or diag_12 like 'M3[24]%' or
                diag_13 like 'M3[24]%' or diag_14 like 'M3[24]%'  then 1
        When    diag_01 = 'M332' or diag_02 = 'M332' or diag_03= 'M332' or diag_04 = 'M332' or diag_05
= 'M332' or diag_06 = 'M332' or
                diag_07 = 'M332'or diag_08 = 'M332' or diag_09 = 'M332' or diag_10 = 'M332' or diag_11
= 'M332' or diag_12 = 'M332' or
                diag_13 = 'M332'or diag_14 = 'M332'       then 1
        When    diag_01 = 'M353' or diag_02 = 'M353' or diag_03= 'M353' or diag_04 = 'M353' or diag_05
= 'M353' or diag_06 = 'M353' or
                diag_07 = 'M353'or diag_08 = 'M353' or diag_09 = 'M353' or diag_10 = 'M353' or diag_11
= 'M353' or diag_12 = 'M353' or
```

```
                diag_13 = 'M353'or diag_14 = 'M353'  then 1              else 0 end ;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Dementia = case when diag_01 like 'F0[0123]%' or diag_02 like 'F0[0123]%' or diag_03 like
'F0[0123]%'  or diag_04 like 'F0[0123]%'  or diag_05 like 'F0[0123]%'
                 or diag_06 like 'F0[0123]%'  or diag_07 like 'F0[0123]%'  or diag_08 like 'F0[0123]%'
or diag_09 like 'F0[0123]%'  or diag_10 like 'F0[0123]%'
                  or diag_11 like 'F0[0123]%'  or diag_12 like 'F0[0123]%'  or diag_13 like
'F0[0123]%'  or diag_14 like 'F0[0123]%'  then 1
        WHEN  diag_01 like 'F051%' or diag_02 like 'F051%' or diag_03 like 'F051%' or diag_04 like
'F051%' or diag_05 like 'F051%' or diag_06 like 'F051%' or
                diag_07 like 'F051%' or diag_08 like 'F051%' or diag_09 like 'F051%' or diag_10 like
'F051%' or diag_11 like 'F051%' or diag_12 like 'F051%' or
                diag_13 like 'F051%' or diag_14 like 'F051%'      then 1 else 0 end ;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diabetes = case when diag_01 like 'E1[0134][15689]%' or diag_02 like 'E1[0134][15689]%' or diag_03
like 'E1[0134][15689]%'  or diag_04 like 'E1[0134][15689]%'  or diag_05 like 'E1[0134][15689]%'
                 or diag_06 like 'E1[0134][15689]%'  or diag_07 like 'E1[0134][15689]%'  or diag_08
like 'E1[0134][15689]%'  or diag_09 like 'E1[0134][15689]%'  or diag_10 like 'E1[0134][15689]%'
                  or diag_11 like 'E1[0134][15689]%'  or diag_12 like 'E1[0134][15689]%'  or diag_13
like 'E1[0134][15689]%'  or diag_14 like 'E1[0134][15689]%'  then 1 else 0 end;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Liver_disease =  case when diag_01 like 'K70[23]%' or diag_02 like 'K70[23]%' or diag_03 like
'K70[23]%'  or diag_04 like 'K70[23]%'  or diag_05 like 'K70[23]%'
                 or diag_06 like 'K70[23]%'  or diag_07 like 'K70[23]%'  or diag_08 like 'K70[23]%'
or diag_09 like 'K70[23]%'  or diag_10 like 'K70[23]%'
                  or diag_11 like 'K70[23]%'  or diag_12 like 'K70[23]%'  or diag_13 like 'K70[23]%'
or diag_14 like 'K70[23]%'    then 1
        WHEN  diag_01 like 'K717%' or diag_02 like 'K717%' or diag_03 like 'K717%' or diag_04 like
'K717%' or diag_05 like 'K717%' or diag_06 like 'K717%' or
                diag_07 like 'K717%' or diag_08 like 'K717%' or diag_09 like 'K717%' or diag_10 like
'K717%' or diag_11 like 'K717%' or diag_12 like 'K717%' or
                diag_13 like 'K717%' or diag_14 like 'K717%'      then 1
        when diag_01 like 'K7[34]%' or diag_02 like 'K7[34]%' or diag_03 like 'K7[34]%'  or diag_04
like 'K7[34]%'  or diag_05 like 'K7[34]%'
                 or diag_06 like 'K7[34]%'  or diag_07 like 'K7[34]%'  or diag_08 like 'K7[34]%'  or
diag_09 like 'K7[34]%'  or diag_10 like 'K7[34]%'
                  or diag_11 like 'K7[34]%'  or diag_12 like 'K7[34]%'  or diag_13 like 'K7[34]%'  or
diag_14 like 'K7[34]%' then 1 else 0 end ;

-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Peptic_Ulcer = case when diag_01 like 'k2[5678]%' or diag_02 like 'k2[5678]%' or diag_03 like
'k2[5678]%'  or diag_04 like 'k2[5678]%'  or diag_05 like 'k2[5678]%'
                 or diag_06 like 'k2[5678]%'  or diag_07 like 'k2[5678]%'  or diag_08 like 'k2[5678]%'
or diag_09 like 'k2[5678]%'  or diag_10 like 'k2[5678]%'
                  or diag_11 like 'k2[5678]%'  or diag_12 like 'k2[5678]%'  or diag_13 like
'k2[5678]%'  or diag_14 like 'k2[5678]%'  then 1 else 0 end;

-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Peripheral_vasc_disease = case when diag_01 in ('I739','I790','z958', 'z959') or diag_02 in
('I739','I790','z958', 'z959')or diag_03 in ('I739','I790','z958', 'z959')  or diag_04 in
('I739','I790','z958', 'z959') or diag_05 in ('I739','I790','z958', 'z959')
                 or diag_06 in ('I739','I790','z958', 'z959')  or diag_07 in ('I739','I790','z958',
'z959')  or diag_08 in ('I739','I790','z958', 'z959')  or diag_09 in ('I739','I790','z958', 'z959')
or diag_10 in ('I739','I790','z958', 'z959')
                  or diag_11 in ('I739','I790','z958', 'z959')  or diag_12 in ('I739','I790','z958',
'z959')  or diag_13 in ('I739','I790','z958', 'z959') or diag_14 in ('I739','I790','z958', 'z959')
then 1
        WHEN  diag_01 like 'I71%' or diag_02 like 'I71%' or diag_03 like 'I71%' or diag_04 like 'I71%'
or diag_05 like 'I71%' or diag_06 like 'I71%' or
                diag_07 like 'I71%' or diag_08 like 'I71%' or diag_09 like 'I71%' or diag_10 like
'I71%' or diag_11 like 'I71%' or diag_12 like 'I71%' or
                diag_13 like 'I71%' or diag_14 like 'I71%'        then 1
        when diag_01 like 'r02%' or diag_02 like 'r02%' or diag_03 like 'r02%'  or diag_04 like 'r02%'
or diag_05 like 'r02%'
                 or diag_06 like 'r02%'  or diag_07 like 'r02%'  or diag_08 like 'r02%'  or diag_09
like 'r02%'  or diag_10 like 'r02%'
                  or diag_11 like 'r02%'  or diag_12 like 'r02%'  or diag_13 like 'r02%'  or diag_14
like 'r02%' then 1 else 0 end ;
```

```
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Pulmonary_disease = case     WHEN  diag_01 like 'J4[01234567]%' or diag_02 like 'J4[01234567]%' or
diag_03 like 'J4[01234567]%' or diag_04 like 'J4[01234567]%' or diag_05 like 'J4[01234567]%' or
diag_06 like 'J4[01234567]%' or
                diag_07 like 'J4[01234567]%' or diag_08 like 'J4[01234567]%' or diag_09 like
'J4[01234567]%' or diag_10 like 'J4[01234567]%' or diag_11 like 'J4[01234567]%' or diag_12 like
'J4[01234567]%' or
                diag_13 like 'J4[01234567]%' or diag_14 like 'J4[01234567]%'  then 1
        when diag_01 like 'J6[01234567]%' or diag_02 like 'J6[01234567]%' or diag_03 like
'J6[01234567]%'  or diag_04 like 'J6[01234567]%'  or diag_05 like 'J6[01234567]%'
                 or diag_06 like 'J6[01234567]%'  or diag_07 like 'J6[01234567]%'  or diag_08 like
'J6[01234567]%'  or diag_09 like 'J6[01234567]%'  or diag_10 like 'J6[01234567]%'
                  or diag_11 like 'J6[01234567]%'  or diag_12 like 'J6[01234567]%'  or diag_13 like
'J6[01234567]%'  or diag_14 like 'J6[01234567]%'   then 1 else 0 end ;

-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Cancer = case       WHEN  diag_01 like 'c[0345689]%' or diag_02 like 'c[0345689]%' or diag_03 like
'c[0345689]%' or diag_04 like 'c[0345689]%' or diag_05 like 'c[0345689]%' or diag_06 like
'c[0345689]%' or
                diag_07 like 'c[0345689]%' or diag_08 like 'c[0345689]%' or diag_09 like 'c[0345689]%'
or diag_10 like 'c[0345689]%' or diag_11 like 'c[0345689]%' or diag_12 like 'c[0345689]%' or
                diag_13 like 'c[0345689]%' or diag_14 like 'c[0345689]%'  then 1
        WHEN  diag_01 like 'c1[01243567]%' or diag_02 like 'c1[01243567]%' or diag_03 like
'c1[01243567]%' or diag_04 like 'c1[01243567]%' or diag_05 like 'c1[01243567]%' or diag_06 like
'c1[01243567]%' or
                diag_07 like 'c1[01243567]%' or diag_08 like 'c1[01243567]%' or diag_09 like
'c1[01243567]%' or diag_10 like 'c1[01243567]%' or diag_11 like 'c1[01243567]%' or diag_12 like
'c1[01243567]%' or
                diag_13 like 'c1[01243567]%' or diag_14 like 'c1[01243567]%'  then 1
        WHEN  diag_01 like 'c2[12346789]%' or diag_02 like 'c2[12346789]%' or diag_03 like
'c2[12346789]%' or diag_04 like 'c2[12346789]%' or diag_05 like 'c2[12346789]%' or diag_06 like
'c2[12346789]%' or
                diag_07 like 'c2[12346789]%' or diag_08 like 'c2[12346789]%' or diag_09 like
'c2[12346789]%' or diag_10 like 'c2[12346789]%' or diag_11 like 'c2[12346789]%' or diag_12 like
'c2[12346789]%' or
                diag_13 like 'c2[12346789]%' or diag_14 like 'c2[12346789]%' then 1
        when diag_01 like 'c7[0123456]%' or diag_02 like 'c7[0123456]%' or diag_03 like 'c7[0123456]%'
or diag_04 like 'c7[0123456]%'  or diag_05 like 'c7[0123456]%'
                 or diag_06 like 'c7[0123456]%'  or diag_07 like 'c7[0123456]%'  or diag_08 like
'c7[0123456]%'  or diag_09 like 'c7[0123456]%'  or diag_10 like 'c7[0123456]%'
                  or diag_11 like 'c7[0123456]%'  or diag_12 like 'c7[0123456]%'  or diag_13 like
'c7[0123456]%'  or diag_14 like 'c7[0123456]%'   then 1 else 0 end;

-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diabetes_complications =  case when diag_01 like 'E1[0134][2347]%' or diag_02 like
'E1[0134][2347]%' or diag_03 like 'E1[0134][2347]%'  or diag_04 like 'E1[0134][2347]%'  or diag_05
like 'E1[0134][2347]%'
                 or diag_06 like 'E1[0134][2347]%'  or diag_07 like 'E1[0134][2347]%'  or diag_08 like
'E1[0134][2347]%'  or diag_09 like 'E1[0134][2347]%'  or diag_10 like 'E1[0134][2347]%'
                  or diag_11 like 'E1[0134][2347]%'  or diag_12 like 'E1[0134][2347]%'  or diag_13
like 'E1[0134][2347]%'  or diag_14 like 'E1[0134][2347]%'    then 1 else 0 end ;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Paraplegia =case when diag_01 in ('G041',  'G820', 'G821', 'G822') or diag_02 in ('G041',  'G820',
'G821', 'G822')or diag_03 in ('G041',  'G820', 'G821', 'G822')  or diag_04 in ('G041',  'G820',
'G821', 'G822') or diag_05 in ('G041',  'G820', 'G821', 'G822')
                 or diag_06 in ('G041',  'G820', 'G821', 'G822')  or diag_07 in ('G041',  'G820',
'G821', 'G822')  or diag_08 in ('G041',  'G820', 'G821', 'G822')  or diag_09 in ('G041',  'G820',
'G821', 'G822')  or diag_10 in ('G041',  'G820', 'G821', 'G822')
                  or diag_11 in ('G041',  'G820', 'G821', 'G822')  or diag_12 in ('G041',  'G820',
'G821', 'G822')  or diag_13 in ('G041',  'G820', 'G821', 'G822') or diag_14 in ('G041',  'G820',
'G821', 'G822')  then 1
        WHEN  diag_01 like 'G81%' or diag_02 like 'G81%' or diag_03 like 'G81%' or diag_04 like 'G81%'
or diag_05 like 'G81%' or diag_06 like 'G81%' or
                diag_07 like 'G81%' or diag_08 like 'G81%' or diag_09 like 'G81%' or diag_10 like
'G81%' or diag_11 like 'G81%' or diag_12 like 'G81%' or
                diag_13 like 'G81%' or diag_14 like 'G81%' then 1         else 0 end ;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Metastatic_cancer =case      WHEN  diag_01 like 'c7[789]%' or diag_02 like 'c7[789]%' or diag_03
like 'c7[789]%' or diag_04 like 'c7[789]%' or diag_05 like 'c7[789]%' or diag_06 like 'c7[789]%' or
```

```
               diag_07 like 'c7[789]%' or diag_08 like 'c7[789]%' or diag_09 like 'c7[789]%' or
diag_10 like 'c7[789]%' or diag_11 like 'c7[789]%' or diag_12 like 'c7[789]%' or
               diag_13 like 'c7[789]%' or diag_14 like 'c7[789]%'  then 1        else 0 end ;

-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set HIV=  case  WHEN  diag_01 like 'B2[01234]%' or diag_02 like 'B2[01234]%' or diag_03 like
'B2[01234]%' or diag_04 like 'B2[01234]%' or diag_05 like 'B2[01234]%' or diag_06 like 'B2[01234]%' or
               diag_07 like 'B2[01234]%' or diag_08 like 'B2[01234]%' or diag_09 like 'B2[01234]%' or
diag_10 like 'B2[01234]%' or diag_11 like 'B2[01234]%' or diag_12 like 'B2[01234]%' or
               diag_13 like 'B2[01234]%' or diag_14 like 'B2[01234]%'    then 1  else 0 end ;
-----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Severe_liver_dis=  case when diag_01 in ('k721',  'k729', 'k766', 'k767') or diag_02 in ('k721',
'k729', 'k766', 'k767')or diag_03 in ('k721',  'k729', 'k766', 'k767') or diag_04 in ('k721',  'k729',
'k766', 'k767') or diag_05 in ('k721',  'k729', 'k766', 'k767')
                or diag_06 in ('k721',  'k729', 'k766', 'k767')  or diag_07 in ('k721',  'k729',
'k766', 'k767')  or diag_08 in ('k721',  'k729', 'k766', 'k767')  or diag_09 in ('k721',  'k729',
'k766', 'k767')  or diag_10 in ('k721',  'k729', 'k766', 'k767')
                 or diag_11 in ('k721',  'k729', 'k766', 'k767')  or diag_12 in ('k721',  'k729',
'k766', 'k767')  or diag_13 in ('k721',  'k729', 'k766', 'k767') or diag_14 in ('k721',  'k729',
'k766', 'k767')  then 1 else 0 end ;
-----(84192 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Renal_disease =  case        WHEN  diag_01 like 'I1[23]%' or diag_02 like 'I1[23]%' or diag_03
like 'I1[23]%' or diag_04 like 'I1[23]%' or diag_05 like 'I1[23]%' or diag_06 like 'I1[23]%' or
               diag_07 like 'I1[23]%' or diag_08 like 'I1[23]%' or diag_09 like 'I1[23]%' or diag_10
like 'I1[23]%' or diag_11 like 'I1[23]%' or diag_12 like 'I1[23]%' or
               diag_13 like 'I1[23]%' or diag_14 like 'I1[23]%'  then 1
        when diag_01 like 'N0[13]%' or diag_02 like 'N0[13]%' or diag_03 like 'N0[13]%'  or diag_04
like 'N0[13]%'  or diag_05 like 'N0[13]%'
                 or diag_06 like 'N0[13]%'  or diag_07 like 'N0[13]%'  or diag_08 like 'N0[13]%'  or
diag_09 like 'N0[13]%'  or diag_10 like 'N0[13]%'
                  or diag_11 like 'N0[13]%'  or diag_12 like 'N0[13]%'  or diag_13 like 'N0[13]%'  or
diag_14 like 'N0[13]%'    then 1
        when diag_01 like 'N1[89]%' or diag_02 like 'N1[89]%' or diag_03 like 'N1[89]%'  or diag_04
like 'N1[89]%'  or diag_05 like 'N1[89]%'
                 or diag_06 like 'N1[89]%'  or diag_07 like 'N1[89]%'  or diag_08 like 'N1[89]%'  or
diag_09 like 'N1[89]%'  or diag_10 like 'N1[89]%'
                  or diag_11 like 'N1[89]%'  or diag_12 like 'N1[89]%'  or diag_13 like 'N1[89]%'  or
diag_14 like 'N1[89]%'    then 1
         when diag_01 like 'N25%' or diag_02 like 'N25%' or diag_03 like 'N25%'  or diag_04 like
'N25%'  or diag_05 like 'N25%'
                or diag_06 like 'N25%'  or diag_07 like 'N25%'  or diag_08 like 'N25%'  or diag_09
like 'N25%'  or diag_10 like 'N25%'
                 or diag_11 like 'N25%'  or diag_12 like 'N25%'  or diag_13 like 'N25%'  or diag_14
like 'N25%'  then 1
        when diag_01 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074') or diag_02 in
('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')or diag_03 in ('N052', 'N053', 'n054',
'N055', 'N056', 'N072', 'N073', 'N074')  or diag_04 in ('N052', 'N053', 'n054', 'N055', 'N056',
'N072', 'N073', 'N074') or diag_05 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')
                 or diag_06 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or
diag_07 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or diag_08 in ('N052',
'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or diag_09 in ('N052', 'N053', 'n054',
'N055', 'N056', 'N072', 'N073', 'N074')  or diag_10 in ('N052', 'N053', 'n054', 'N055', 'N056',
'N072', 'N073', 'N074')
                 or diag_11 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or
diag_12 in ('N052', 'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074')  or diag_13 in ('N052',
'N053', 'n054', 'N055', 'N056', 'N072', 'N073', 'N074') or diag_14 in ('N052', 'N053', 'n054', 'N055',
'N056', 'N072', 'N073', 'N074')  then 1
                  else 0 end;
-----(84192 row(s) affected)



---there are some missing colonoscopy dates still
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set colonoscopy_date = DISDATE
where EPISTART = DISDATE and colonoscopy_date is null
---52
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set colonoscopy_date = Epistart
where  colonoscopy_date is null
```

```
---(35 row(s) affected)

select droprecord, admidate, disdate, epistart, colonoscopy_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
order by colonoscopy_date

---update those records where the endoscopy date is outside the period we are looking at
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where  colonoscopy_date  <  '2003/01/04'
----(0 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where  colonoscopy_date  >  '2009/31/03'
----24
---Need to add ethnicity to these tables
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set ethnos = t2.ethnos
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                                                    on t1.Patient_ID = t2.Patient_ID
                                                                    and t1.SPELL_ID = t2.SPELL_ID
                                                                    and t1.EPIORDER = t2.EPIORDER

-----(32708 row(s) affected)

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set ethnos = t2.ethnos
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                                                    on t1.Patient_ID = t2.Patient_ID
                                                                    and t1.SPELL_ID = t2.SPELL_ID
                                                                    and t1.EPIORDER = t2.EPIORDER
---(51484 row(s) affected)

---lets look at these ethnicities
select ethnos, COUNT(*)
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by ethnos
order by ethnos
---there are quite  a  lot here which are invalid
---lets have a look whether there are any better ethnicities and any patietns who only have one
ethnicity recorded
select  t1.patient_ID, COUNT(distinct t2.ethnos)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1  inner join (select Patient_ID,
Ethnos, admidate

                                                        from HED.dbo.HES_EPISODES_Part1

                                                        where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')

                                                        union all

                                                        select Patient_ID, Ethnos , admidate

                                                        from HED.dbo.HES_EPISODES

                                                        where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2

        on t1.Patient_ID = t2.Patient_ID
where t1.ethnos  in ('0','1','2','3','4','5','6','7','8','9','x','y','z')
group by t1.Patient_ID
having COUNT(distinct t2.ethnos) =1
----9655 rows
---updating these patietns
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set ethnos = t2.ETHNOS
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Ethnos, admidate

                                                        from HED.dbo.HES_EPISODES_Part1
```

```
                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')
                                                        and Patient_ID in (select
t1.patient_ID
        from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select
Patient_ID, Ethnos, admidate
                                                    from HED.dbo.HES_EPISODES_Part1
                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')
                                                    union all
                                                    select Patient_ID, Ethnos , admidate
                                                    from HED.dbo.HES_EPISODES
                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2
                on t1.Patient_ID = t2.Patient_ID
        where t1.ethnos  in ('0','1','2','3','4','5','6','7','8','9','x','y','z')
        group by t1.Patient_ID
        having COUNT(distinct t2.ethnos) =1 )
                                                    union all
                                                    select Patient_ID, Ethnos , admidate
                                                    from HED.dbo.HES_EPISODES
                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')
                                                        and Patient_ID in (select
t1.patient_ID
        from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select
Patient_ID, Ethnos, admidate
                                                    from HED.dbo.HES_EPISODES_Part1
                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')
```

```
                                                        union all

                                                        select Patient_ID, Ethnos , admidate

                                                        from HED.dbo.HES_EPISODES

                                                        where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2


                on t1.Patient_ID = t2.Patient_ID

        where t1.ethnos  in ('0','1','2','3','4','5','6','7','8','9','x','y','z')

        group by t1.Patient_ID

        having COUNT(distinct t2.ethnos) =1 )) t2

                on t1.Patient_ID = t2.PAtient_ID
---(11318 row(s) affected)
---having a look at the patients now, who have a proper ethnicity, but more than one recorded
select t1.Patient_ID, t2.ethnos, COUNT(distinct t2.ADMIDATE) admissions
into #ethnicmastectomy
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Ethnos, admidate

                                                        from HED.dbo.HES_EPISODES_Part1

                                                        where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')

                                                        union all

                                                        select Patient_ID, Ethnos , admidate

                                                        from HED.dbo.HES_EPISODES

                                                        where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2

        on t1.Patient_ID = t2.Patient_ID
where  t1.patient_ID in (select  t1.patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID, Ethnos,
admidate


        from HED.dbo.HES_EPISODES_Part1

        where ETHNOS not in ('0','1','2','3','4','5','6','7','8','9','x','y','z')

        union all

        select Patient_ID, Ethnos , admidate

        from HED.dbo.HES_EPISODES
```

```
        where ETHNOS not in ('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2

                                                    on t1.Patient_ID = t2.Patient_ID
                                                where t1.ethnos  in
('0','1','2','3','4','5','6','7','8','9','x','y','z')
                                                group by t1.Patient_ID
                                                having COUNT(distinct t2.ethnos) >1 )
group by t1.Patient_ID, t2.ethnos
-----(1544 row(s) affected

select distinct PAtient_ID
from #ethnicmastectomy;
----742
select t1.*
from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions) maxadmi
                                                                                    from
#ethnicmastectomy

        group by Patient_ID) t2
                                                                on t1.Patient_ID = t2.Patient_ID
                                                                and t1.admissions = t2.maxadmi
---812

select Patient_ID, COUNT(distinct ethnos) ethnicities
into #ethnicUniquePatient
from (select t1.*
                from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions) maxadmi

                from  #ethnicmastectomy

                group by Patient_ID) t2
                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                        and t1.admissions =
t2.maxadmi) a
group by Patient_ID
having COUNT(distinct ethnos) = 1
----673

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set ethnos = t2.ethnos
from   Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1    inner join  (select t1.*

                        from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions)
maxadmi

                                                from  #ethnicmastectomy


                                                group by Patient_ID) t2


                on t1.Patient_ID = t2.Patient_ID


                and t1.admissions = t2.maxadmi

                        where t1.Patient_ID in (select distinct Patient_ID

                                                                from #ethnicUniquePatient) ) t2
                                                                        on
t1.PAtient_ID = t2.Patient_ID
where t1.ethnos in ('0','1','2','3','4','5','6','7','8','9','x','y','z')

---745

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set ethnos = t2.ethnos
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select t1.* ,
t4.lastadmission
```

```
                from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions) maxadmi
                                        from  #ethnicmastectomy
                                        group by Patient_ID) t2
        on t1.Patient_ID = t2.Patient_ID
        and t1.admissions = t2.maxadmi
                                                                    inner join (select
Patient_ID, Ethnos, MAX(admidate)lastadmission
        from (select Patient_ID, admidate , Ethnos
                from HED.dbo.HES_Episodes
                union all
                select Patient_ID, admidate, Ethnos
                from HED.dbo.HES_EPISODES_Part1) a
        group by Patient_ID, Ethnos) t3
                                                                                on
t1.Patient_ID = t3.Patient_ID
                                                                                and
t1.ETHNOS = t3.ETHNOS
                                                                    inner join (select
Patient_ID, MAX(admidate)lastadmission
        from (select Patient_ID, admidate , Ethnos
                from HED.dbo.HES_Episodes
                union all
                select Patient_ID, admidate, Ethnos
                from HED.dbo.HES_EPISODES_Part1) a
        group by Patient_ID) t4
                                                                            on
t1.Patient_ID = t4.Patient_ID
                                                                            and
t3.lastadmission = t4.lastadmission
                where t1.Patient_ID not in (select distinct Patient_ID
                                                                                from
#ethnicUniquePatient) ) t2
```

```
                                                                                on
t1.Patient_ID = t2.Patient_ID
where t1.ethnos in ('0','1','2','3','4','5','6','7','8','9','x','y','z')
        ----    (68 row(s) affected)


---lets check out the ethnicites again
select ethnos, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
group by ethnos
order by ethnos
---this is a huge improvement on the previous data - still some old codes though...

alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add  ethnic_group varchar(50);

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set ethnic_group = case     when ethnos in ('A','B','C','0') then 'White'
                                                when ethnos in ('D','E','F','G') then 'Mixed'
                                                when ethnos in ('H','J','K','L','4','5','6') then
'Asian or Asian British'
                                                when ethnos in ('M','N','P','1','2','3') then 'Black
or Black British'
                                                when ethnos in ('R','7') then 'Chinese'
                                                when ethnos in ('S','8') then 'Other Ethnic Groups'
                                                when ethnos in ('Z','X','9') then 'Unknown' end
----84192


alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add  Charlson varchar(2);

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Charlson = t2.DOMI_0214_CHARLSON_NWT
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_SPELLS t2
                                                on t1.PAtient_ID = t2.Patient_ID
                                                and t1.Spell_ID = t2.Spell_ID
        ---(84099 row(s) affected)


select top 100 *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add site_Colonoscopy varchar(5);

select
case when opertn_01 IN ('H221','H228','H229','H201', 'H206','H208') and( opertn_02 like 'Z2%'  or
opertn_02 like 'O[13]%') then Opertn_02
        when opertn_01 IN ('H221','H228','H229','H201' ,'H206','H208')   and opertn_02 not like 'G%'
and (opertn_03 like 'Z2%' or opertn_03 like 'O[13]%')then Opertn_03
        when opertn_01 IN ('H221','H228','H229','H201' ,'H206','H208') and opertn_02 not like 'G%' and
opertn_03 not like 'G%' and (opertn_04 like 'Z2%'or opertn_04 like 'O[13]%') then Opertn_04
        when opertn_02 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_03 like 'Z2%'  or
opertn_03 like 'O[13]%') then Opertn_03
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_04 like 'Z2%'or
opertn_04 like 'O[13]%') then Opertn_04
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_04 not like 'G%' and
(opertn_05 like 'Z2%'or opertn_05 like 'O[13]%') then Opertn_05
        when opertn_04 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_05 like 'Z2%'or
opertn_05 like 'O[13]%') then Opertn_05
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_04 not like 'G%' and
(opertn_05 like 'Z2%'or opertn_05 like 'O[13]%') then Opertn_05
        when opertn_05 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_06 like 'Z2%'or
opertn_06 like 'O[13]%') then Opertn_06
        when opertn_05 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_06 not like 'G%' and
(opertn_07 like 'Z2%'or opertn_07 like 'O[13]%') then Opertn_07
        when opertn_06 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_07 like 'Z2%'or
opertn_07 like 'O[13]%') then Opertn_07
        when opertn_06 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_07 not like 'G%' and
(opertn_08 like 'Z2%'or opertn_08 like 'O[13]%') then Opertn_08
        when opertn_07 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_08 like 'Z2%'or
opertn_08 like 'O[13]%') then Opertn_08
```

```
        when opertn_07 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_08 not like 'G%' and
(opertn_09 like 'Z2%'or opertn_09 like 'O[13]%') then Opertn_09
        when opertn_08 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_09 like 'Z2%'or
opertn_09 like 'O[13]%' )then Opertn_09
        when opertn_08 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_09 not like 'G%' and
(opertn_10 like 'Z2%'or opertn_10 like 'O[13]%') then Opertn_10
        when opertn_09 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_10 like 'Z2%'or
opertn_10 like 'O[13]%') then Opertn_10
        when opertn_09 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_10 not like 'G%' and
(opertn_11 like 'Z2%'or opertn_11 like 'O[13]%') then Opertn_11
        when opertn_10 IN ('H221','H228','H229','H201', 'H206','H208') and( opertn_11 like 'Z2%'or
opertn_11 like 'O[13]%' )then Opertn_11
        when opertn_10 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_11 not like 'G%' and
(opertn_12 like 'Z2%'or opertn_12 like 'O[13]%' )then Opertn_12
        when opertn_11 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_12 like 'Z2%'or
opertn_12 like 'O[13]%') then Opertn_12
         else null end , *
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
order by case when opertn_01 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_02 like 'Z2%'
or opertn_02 like 'O[13]%') then Opertn_02
        when opertn_01 IN ('H221','H228','H229','H201' ,'H206','H208')   and opertn_02 not like 'G%'
and (opertn_03 like 'Z2%' or opertn_03 like 'O[13]%')then Opertn_03
        when opertn_01 IN ('H221','H228','H229','H201' ,'H206','H208') and opertn_02 not like 'G%' and
opertn_03 not like 'G%' and (opertn_04 like 'Z2%'or opertn_04 like 'O[13]%') then Opertn_04
        when opertn_02 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_03 like 'Z2%'  or
opertn_03 like 'O[13]%') then Opertn_03
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_04 like 'Z2%'or
opertn_04 like 'O[13]%') then Opertn_04
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_04 not like 'G%' and
(opertn_05 like 'Z2%'or opertn_05 like 'O[13]%') then Opertn_05
        when opertn_04 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_05 like 'Z2%'or
opertn_05 like 'O[13]%') then Opertn_05
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_04 not like 'G%' and
(opertn_05 like 'Z2%'or opertn_05 like 'O[13]%') then Opertn_05
        when opertn_05 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_06 like 'Z2%'or
opertn_06 like 'O[13]%') then Opertn_06
        when opertn_05 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_06 not like 'G%' and
(opertn_07 like 'Z2%'or opertn_07 like 'O[13]%') then Opertn_07
        when opertn_06 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_07 like 'Z2%'or
opertn_07 like 'O[13]%') then Opertn_07
        when opertn_06 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_07 not like 'G%' and
(opertn_08 like 'Z2%'or opertn_08 like 'O[13]%') then Opertn_08
        when opertn_07 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_08 like 'Z2%'or
opertn_08 like 'O[13]%') then Opertn_08
        when opertn_07 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_08 not like 'G%' and
(opertn_09 like 'Z2%'or opertn_09 like 'O[13]%') then Opertn_09
        when opertn_08 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_09 like 'Z2%'or
opertn_09 like 'O[13]%' )then Opertn_09
        when opertn_08 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_09 not like 'G%' and
(opertn_10 like 'Z2%'or opertn_10 like 'O[13]%') then Opertn_10
        when opertn_09 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_10 like 'Z2%'or
opertn_10 like 'O[13]%') then Opertn_10
        when opertn_09 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_10 not like 'G%' and
(opertn_11 like 'Z2%'or opertn_11 like 'O[13]%') then Opertn_11
        when opertn_10 IN ('H221','H228','H229','H201', 'H206','H208') and( opertn_11 like 'Z2%'or
opertn_11 like 'O[13]%' )then Opertn_11
        when opertn_10 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_11 not like 'G%' and
(opertn_12 like 'Z2%'or opertn_12 like 'O[13]%' )then Opertn_12
        when opertn_11 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_12 like 'Z2%'or
opertn_12 like 'O[13]%') then Opertn_12
         else null end;
---and now to update
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set site_Colonoscopy = case when opertn_01 IN ('H221','H228','H229','H201', 'H206','H208') and
(opertn_02 like 'Z2%'  or opertn_02 like 'O[13]%') then Opertn_02
        when opertn_01 IN ('H221','H228','H229','H201' ,'H206','H208')   and opertn_02 not like 'G%'
and (opertn_03 like 'Z2%' or opertn_03 like 'O[13]%')then Opertn_03
        when opertn_01 IN ('H221','H228','H229','H201' ,'H206','H208') and opertn_02 not like 'G%' and
opertn_03 not like 'G%' and (opertn_04 like 'Z2%'or opertn_04 like 'O[13]%') then Opertn_04
        when opertn_02 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_03 like 'Z2%'  or
opertn_03 like 'O[13]%') then Opertn_03
        when opertn_02 IN ('H221','H228','H229','H201' ,'H206','H208') and opertn_03 not like 'G%' and
opertn_04 not like 'G%' and (opertn_05 like 'Z2%'or opertn_05 like 'O[13]%') then Opertn_05
```

```
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_04 like 'Z2%'or
opertn_04 like 'O[13]%') then Opertn_04
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_04 not like 'G%' and
(opertn_05 like 'Z2%'or opertn_05 like 'O[13]%') then Opertn_05
        when opertn_04 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_05 like 'Z2%'or
opertn_05 like 'O[13]%') then Opertn_05
        when opertn_03 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_04 not like 'G%' and
(opertn_05 like 'Z2%'or opertn_05 like 'O[13]%') then Opertn_05
        when opertn_05 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_06 like 'Z2%'or
opertn_06 like 'O[13]%') then Opertn_06
        when opertn_05 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_06 not like 'G%' and
(opertn_07 like 'Z2%'or opertn_07 like 'O[13]%') then Opertn_07
        when opertn_06 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_07 like 'Z2%'or
opertn_07 like 'O[13]%') then Opertn_07
        when opertn_06 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_07 not like 'G%' and
(opertn_08 like 'Z2%'or opertn_08 like 'O[13]%') then Opertn_08
        when opertn_07 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_08 like 'Z2%'or
opertn_08 like 'O[13]%') then Opertn_08
        when opertn_07 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_08 not like 'G%' and
(opertn_09 like 'Z2%'or opertn_09 like 'O[13]%') then Opertn_09
        when opertn_08 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_09 like 'Z2%'or
opertn_09 like 'O[13]%' )then Opertn_09
        when opertn_08 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_09 not like 'G%' and
(opertn_10 like 'Z2%'or opertn_10 like 'O[13]%') then Opertn_10
        when opertn_09 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_10 like 'Z2%'or
opertn_10 like 'O[13]%') then Opertn_10
        when opertn_09 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_10 not like 'G%' and
(opertn_11 like 'Z2%'or opertn_11 like 'O[13]%') then Opertn_11
        when opertn_10 IN ('H221','H228','H229','H201', 'H206','H208') and( opertn_11 like 'Z2%'or
opertn_11 like 'O[13]%' )then Opertn_11
        when opertn_10 IN ('H221','H228','H229','H201', 'H206','H208') and opertn_11 not like 'G%' and
(opertn_12 like 'Z2%'or opertn_12 like 'O[13]%' )then Opertn_12
        when opertn_11 IN ('H221','H228','H229','H201', 'H206','H208') and (opertn_12 like 'Z2%'or
opertn_12 like 'O[13]%') then Opertn_12
        when OPERTN_02 like 'Z2[89]%' then OPERTN_02
        when OPERTN_03 like 'Z2[89]%' then OPERTN_03
        when OPERTN_04 like 'Z2[89]%' then OPERTN_04
        when OPERTN_05 like 'Z2[89]%' then OPERTN_05
        when OPERTN_06 like 'Z2[89]%' then OPERTN_06
        when OPERTN_07 like 'Z2[89]%' then OPERTN_07
        when OPERTN_08 like 'Z2[89]%' then OPERTN_08
        when OPERTN_09 like 'Z2[89]%' then OPERTN_09
        when OPERTN_10 like 'Z2[89]%' then OPERTN_10
         else null end
-----(84192 row(s) affected)



SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                    from HES_DW.dbo.HES_OP_200607

                                    where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                    on
t1.Patient_ID = t2.PatientID
```

```
                                                                                and
t1.firstAdmissionDiagnosis <= t2.Apptdate
------  (505 row(s) affected)


SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200708

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                on
t1.Patient_ID = t2.PatientID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.Apptdate
----(770 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200809

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                on
t1.Patient_ID = t2.PatientID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.Apptdate

----(1343 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200910

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'
```

```
                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate
----(1133 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201011

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate

----(574 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HEs_DW.dbo.HES_OP_201112

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate
----(514 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
```

```
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HEs_DW.dbo.HES_OP_201213

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.Apptdate
---     (484 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HEs_DW.dbo.HES_OP_201314

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%x7[0123]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X352%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X384%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%X373%') t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                        and
t1.firstAdmissionDiagnosis <= t2.Apptdate
---     (389 row(s) affected)
---and updating chemo using diagnosis codes


SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200607

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z082%'

                                                or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z511%'

                                                or
```

```
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z542%') t2
                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate

----(0 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200708

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z082%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z511%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z542%') t2
                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate
----(194 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200809

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z082%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z511%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z542%') t2
                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate

----    (311 row(s) affected)           /* updated to here */

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
Apptdate

                                        from HES_DW.dbo.HES_OP_200910

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z082%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z511%'
```

```
                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z542%') t2
                                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate
-----(238 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201011

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z082%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z511%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z542%') t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate
---     (125 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201112

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z082%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z511%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z542%') t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate
----(68 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set chemo = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
Apptdate

                                        from HES_DW.dbo.HES_OP_201213

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z082%'

                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z511%'
```

```
                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z542%') t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.Apptdate
---(70 row(s) affected)
-----updating radiotherapy
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                        from HED.dbo.HES_episodes

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.epistart
----    (1347 row(s) affected
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                        from HED.dbo.HES_episodes_part1

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.epistart
----(97 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                        from HES_DW.dbo.HES_OP_200607

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.apptdate
----(149 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
```

```
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                        from HES_DW.dbo.HES_OP_200708

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.apptdate
----(239 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                        from HES_DW.dbo.HES_OP_200809

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.apptdate

---(420row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                        from HES_DW.dbo.HES_OP_200910

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                                    on
t1.Patient_ID = t2.PatientID
                                                                                                    and
t1.firstAdmissionDiagnosis <= t2.apptdate
---(356 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
apptdate

                                        from HES_DW.dbo.HES_OP_201011

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                              or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
```

```
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----(287 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
apptdate

                                        from HES_DW.dbo.HES_OP_201112

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----    (228 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
apptdate

                                        from HES_DW.dbo.HES_OP_201213

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%x6[34578]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+opertn_10+op
ertn_11+opertn_12 like '%Y9[12]%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----(706 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
admidate

                                        from HED.dbo.HES_episodes_part1

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z081%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z510%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z541%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.Admidate

-----(706 row(s) affected)
```

```
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                    from HED.dbo.HES_episodes

                                    where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z081%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z510%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z541%') t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.epistart


----(1312 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
epistart

                                    from HED.dbo.HES_episodes_part1

                                    where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z081%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z510%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08
+DIAG_09+DIAG_10+DIAG_11+DIAG_12+DIAG_13+DIAG_14+DIAG_15 like '%z541%') t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.epistart
----            (708 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                    from HES_DW.dbo.HES_OP_200607

                                    where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z081%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z510%'

                                                        or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z541%') t2
                                                                                    on
t1.Patient_ID = t2.PatientID
```

```
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----(0 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                        from HES_DW.dbo.HES_OP_200708

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z081%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z510%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z541%') t2
                                                                                on
t1.Patient_ID = t2.PatientID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
-----(12 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                        from HES_DW.dbo.HES_OP_200809

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z081%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z510%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z541%') t2
                                                                                on
t1.Patient_ID = t2.PatientID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----    (52 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select PatientID,
apptdate

                                        from HES_DW.dbo.HES_OP_200910

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z081%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z510%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z541%') t2
                                                                                on
t1.Patient_ID = t2.PatientID
```

```
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----    (31 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
apptdate

                                        from HES_DW.dbo.HES_OP_201011

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z081%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z510%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z541%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----20 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
apptdate

                                        from HES_DW.dbo.HES_op_201112

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z081%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z510%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z541%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.apptdate
----(19 row(s) affected)
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set radiotherapy = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
apptdate

                                        from HES_DW.dbo.HES_op_201213

                                        where
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z081%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z510%'

                                                                    or
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+DIAG_06+DIAG_07+DIAG_08 +DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%z541%') t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
```

```
                                                
```

```
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%K5[01]%' ;
---(2806 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set rectalBleed  = 1
where
DIAG_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+DIAG_14 like '%K922%' ;
---(273 row(s) affected)


----Nigel and Danny have asjed that I include something about endoscopic procedures
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add endoscopic_Procedures int;

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set endoscopic_Procedures = 0 ;
--the first part og the hes episodes
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set endoscopic_Procedures = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
admidate

                                        from HED.dbo.HES_episodes_part1

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H23[1345689]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H26[12356789]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H244%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H274%'

                                        ) t2
                                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                                        and
t1.firstAdmissionDiagnosis <= t2.Admidate
-----836
---and now the second...
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set endoscopic_Procedures = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
admidate

                                        from HED.dbo.HES_episodes

                                        where
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H23[1345689]%'

                                                or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H26[12356789]%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H244%'

                                                        or
opertn_01+opertn_02+opertn_03+opertn_04+opertn_04+opertn_05+opertn_06+opertn_07+opertn_08+opertn_09+op
ertn_10+opertn_11+opertn_12 like '%H274%'

                                        ) t2
```

```
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.firstAdmissionDiagnosis <= t2.Admidate
----(2165 row(s) affected)
---we're now inter3ested inthe side of the cancer
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add side_cancer varchar(10);
---first investigsating side of the cancer i.e. left or right, Danny advised about whihch cancers are
considered either left or right
set concat_null_yields_null off
select t1.Patient_ID, case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[0234]%' then 'Right'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[567]%' then 'Left'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Left'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Left' else null end
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_episodes_part1 t2
                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                and
t1.firstAdmissionDiagnosis = t2.epistart
                ---and to update

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer =   side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[01234]%' then 'Right'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[567]%' then 'Left'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Left'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Left' else null end side
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes_part1
t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                and
t1.firstAdmissionDiagnosis = t2.epistart) t2
                                                                on t1.Patient_ID = t2.Patient_ID

        -----    (40114 row(s) affected)
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[01234]%' then 'Right'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[567]%' then 'Left'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Left'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Left' else null end side
```

```
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                                and
t1.firstAdmissionDiagnosis = t2.epistart) t2
                                                                on t1.Patient_ID = t2.Patient_ID

----    (41107 row(s) affected)
---Let have a look at the results to check that everything update ok
select side_cancer, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by side_cancer
-----14308 nulls lets look at these
set concat_null_yields_null off
select
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_episodes_part1 t2
                                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                                and
t1.firstAdmissionDiagnosis = t2.epistart
where t1.droprecord is null and t1.side_cancer is null
----there are a lot where we can't tell what the side is .9

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[01234]%' then 'Right'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[567]%' then 'Left'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Left'
                                                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Left'
                                                         when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D01[01]%' then 'Left' else null end ) side
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                                and
t1.firstAdmissionDiagnosis < t2.epistart
                                        where droprecord is null and side_cancer is null and
(t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2
.diag_10+t2.diag_11+t2.diag_12 like '%C1[89]%' or

        t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.dia
g_09+t2.diag_10+t2.diag_11+t2.diag_12 like '%C20%'
                                                                        or
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D01[01]%'    )
                                                                                group by t1.patient_ID) t2
                                                                on t1.Patient_ID = t2.Patient_ID

        -----(5137 row(s) affected)
        ----1666

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
```

```
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[01234]%' then 'Right'
                                                    when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C18[567]%' then 'Left'
                                                    when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Left'
                                                    when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Left'
                                                     when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D01[01]%' then 'Left' else null end ) side
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes_part1
t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                                    and
t1.firstAdmissionDiagnosis < t2.epistart
                                       where droprecord is null and side_cancer is null and
(t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2
.diag_10+t2.diag_11+t2.diag_12 like '%C1[89]%' or

        t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.dia
g_09+t2.diag_10+t2.diag_11+t2.diag_12 like '%C20%'       or

t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D01[01]%' )
                                                                        group by t1.patient_ID) t2
                                                                on t1.Patient_ID = t2.Patient_ID


-----   (3539 row(s) affected)
----1371
---Lets check this out again
select side_cancer, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by side_cancer
---still 8.7k where unknown
---Lets have a  look to see if there are any records which have the cancer side on the death
certificates

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C18[01234]%' then 'Right'
                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C18[567]%' then 'Left'
                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C19%' then 'Left'
                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C20%' then 'Left'
                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
```

```
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D01[01]%' then 'Left'  else null end ) side
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2

                        on t1.Patient_ID = t2.EXTRACT_HESID collate database_default
                                                                                    and
t1.firstAdmissionDiagnosis < t2.DOD
                                       where droprecord is null and side_cancer is null and
(t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2
.CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CA
USE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE
_OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C1[89]%' or

        t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONAT
AL_3+t2.CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_
6+t2.CAUSE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t
2.CAUSE_OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12
like '%C20%'
                                                                        or
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D01[01]%'             )
                                                                  group by t1.patient_ID) t2
                                                          on t1.Patient_ID = t2.Patient_ID


----(1784 row(s) affected)
---(1336 row(s) affected)
select side_cancer, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by side_cancer
--down to 8.2 k
---Right I need to hink about how to deal with thes missing records /speak to NT/DC
---the other thing I needed to do was add in a column which describes the type of colonoscopy i.e.
whether they took a biopsy ot was a asnare
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add type_colonoscopy varchar(25);
---now to update this

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_colonoscopy = case when
OPERTN_01+OPERTN_02+OPERTN_03+OPERTN_04+OPERTN_05+OPERTN_06+OPERTN_07+opertn_08+OPERTN_09+OPERTN_10+OP
ERTN_11+OPERTN_12 like '%H20[16]%' then 'Polypectomy'
                                                        when
OPERTN_01+OPERTN_02+OPERTN_03+OPERTN_04+OPERTN_05+OPERTN_06+OPERTN_07+opertn_08+OPERTN_09+OPERTN_10+OP
ERTN_11+OPERTN_12 like '%H221%' then 'Biopsy'    else 'Other' end

----(84192 row(s) affected)

---LEts have alook at the numebrs od people who are defitinitely missed/probably missed and have
latent CRC at time of colonoscopy
select case  when ([dayspostopColonoscopyDiag]>180) and ([dayspostopColonoscopyDiag]<=365) then
'Definite_Missed'
        when ([dayspostopColonoscopyDiag]>366) and ([dayspostopColonoscopyDiag]<=1096) then
'Probable_Missed'
        when ([dayspostopColonoscopyDiag]>1096) and ([dayspostopColonoscopyDiag]<=1826) then 'Latent'
        else 'control' end,
      ROUND(cast(sum(t2.admissions) as float)/cast(count(t1.Patient_ID)as float),2)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (Select Patient_ID,
COUNT(*) admissions

                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                where droprecord is null
```

```
                                        group by Patient_ID) t2

                                                                                    on
t1.Patient_ID = t2.Patient_ID
where t1.droprecord is null and LastColonoscopybeforeDiag = 1
group by case    when ([dayspostopColonoscopyDiag]>180) and ([dayspostopColonoscopyDiag]<=365) then
'Definite_Missed'
        when ([dayspostopColonoscopyDiag]>366) and ([dayspostopColonoscopyDiag]<=1096) then
'Probable_Missed'
        when ([dayspostopColonoscopyDiag]>1096) and ([dayspostopColonoscopyDiag]<=1826) then 'Latent'
        else 'control' end




---I've now spoken to dC re the issue of the sides where it is not possible to determning side, he is
happy that we record these as unknowns
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = 'Unknown'
where side_cancer is null

----(8316 row(s) affected)

----Danny has written in the methods, which is something that I think I needed to have done that the
diagnosis date is the same date as the
---last colonoscopy in the previous six months - I need to figure out how much this will affect the
results
----Lets have  a look at the number of patietns who have the same colonoscopy date as the admission
date
select distinct PAtient_ID, firstAdmissionDiagnosis
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where dayspostopColonoscopyDiag = 0
----40610
----how many patients are there in total?
select distinct PAtient_ID
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
----75181
---so about 4/7ths of patients had the colonoscopy on the same date of diagnosis - whcih matches with
what Danny has writted
----Lets just check how many patients there are who had a colonoscopy within the last 6 months befpre
the first admission with colonoscopy
select distinct PAtient_ID
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where dayspostopColonoscopyDiag < 183
----62625
---- so this is now the majority of them , meed to try and work out how much difference there is
between first admission and colonoscopy date
select distinct PAtient_ID, firstAdmissionDiagnosis, max(colonoscopy_date)
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where dayspostopColonoscopyDiag < 183
group by  PAtient_ID, firstAdmissionDiagnosis
--to be fair most of these look fairly similar, although It will be difficult to tell how mcuh
difference there is before we look at them in depth
---Lets put in some altered variables which take into account the differences in the diagnosis and
colonoscopy date

alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add Diag_date_altered datetime;
--and now to update this
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = t2.Minadate
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
PAtient_ID, firstAdmissionDiagnosis, Max(colonoscopy_date)Minadate

                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                        where dayspostopColonoscopyDiag < 183 and droprecord
is null
```

```
                                                group by  PAtient_ID, firstAdmissionDiagnosis)t2
                                            on t1.Patient_ID = t2.Patient_ID
-----    (67569 row(s) affected)
/*
----Lets have a look at those records which have not yet been updated - hopefully this is because they
haven't had a colonoscopy within 6 months prior to record of actual CRC diag
select *
--distinct Patient_ID --16,076
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered is null  and droprecord is null
---22019
---I'll want to tlook at these in more depth late, but 20k shouldn't be too bad...

---Lets have a look see how many p[atients there are that have a diagnosis altered date which isn't
the same as the colonsocopy date
select distinct Patient_ID, firstAdmissionDiagnosis, Diag_date_altered
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered <> firstAdmissionDiagnosis       and droprecord is null
----41321 rows here,  althpough a lot do seem to be within a week of the original date, lets have
alook at those which aren't within one week
 select distinct Patient_ID, firstAdmissionDiagnosis,  Diag_date_altered
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where abs(datediff(dd, Diag_date_altered , firstAdmissionDiagnosis)) > 7          and droprecord is
null
----31104 rows here - which may ne a not of a problem
select *
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered <> firstAdmissionDiagnosis       and droprecord is null and (Diag_date_altered
between ADMIDATE and DISDATE)
order by Patient_ID, admidate
----41493
select *
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered <> firstAdmissionDiagnosis       and droprecord is null and (Diag_date_altered
between ADMIDATE and DISDATE) and(
        firstAdmissionDiagnosis between ADMIDATE and DISDATE)
order by Patient_ID, admidate
---12450 -so about 30 k people have a firstadmission diagnosis which is not during an admission we
have a record of here...
-----
select *
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered <> firstAdmissionDiagnosis       and droprecord is null and(
        firstAdmissionDiagnosis between ADMIDATE and DISDATE)
order by Patient_ID, admidate
----12457 - so there are 6 patients who don't seem to have a colonoscopy date during one of these
admissions
*/
---I'm not sure how well any of the above changes have been working


---lets look at these records and see whether a) how many patietns have a different first admission
date

select distinct PAtient_ID, firstAdmissionDiagnosis, diag_date_altered
into #Colonoscopy_date_diags
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
----(75181 row(s) affected)
---Lets double check how many patietns there actually are without a colonoscopy

select *
from #Colonoscopy_date_diags
where Diag_date_altered is null
---12884
---We will then want to link these records to HES to see whether they were maybe having a colonsocopy
which was superceded with another colorectal procedure
select t1.Patient_ID, t1.Admidate, t1.Epistart, t2.firstAdmissionDiagnosis, t1.DIAG_01, t1.DIAG_02,
t1.DIAG_03, t1.DIAG_04, t1.DIAG_05, t1.diag_06, t1.DIAG_07,  t1.Diag_08, t1.DIAG_09,
t1.DIAG_10, t1.DIAG_11, t1.DIAG_12, t1.OPERTN_01, t1.OPERTN_02, t1.OPERTN_03, t1.OPERTN_04,
t1.OPERTN_05, t1.OPERTN_06, t1.OPERTN_07, t1.OPERTN_08,
t1.OPERTN_09, t1.OPERTN_10, t1.OPERTN_11, t1.opertn_12
from HED.dbo.HES_EPISODES t1 inner join (select *
```

```
                                                                    from
#Colonoscopy_date_diags
                                                                    where
Diag_date_altered is  null) t2
                                on t1.Patient_ID = t2.Patient_ID  collate database_default
                                and DATEDIFF(dd, t1.admidate, t2.firstAdmissionDiagnosis) between -10
and 189
order by t1.Patient_ID, t1.admidate
---there appears to be some patients who are having operations like Hemicolectomy

---I think the following patient might have been incorrectly coded as colorectal as they have C819 for
all records then one is C189, one out of about 15
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID = '0011389D1B58613FF26A141D7C101086'
---Right it looks as though there are some patients who had sigmoidoscopes/hemicolectomies around or
other operations in the right area at the
---time of diagnoisis it may be worthwhile updating these records accordingly to the date of first "H"
 ---procedure - this will mean undoing what I have already doen
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = null
----(84192 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = t2.Minadate
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select distinct
PAtient_ID, firstAdmissionDiagnosis, Max(colonoscopy_date)Minadate

                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                where dayspostopColonoscopyDiag < 5

                                                group by  PAtient_ID, firstAdmissionDiagnosis)t2
                                        on t1.Patient_ID = t2.Patient_ID
----(51138 row(s) affected)
---again lets look at the paients who don't have a record of colonoscopy
select distinct Patient_ID  , firstAdmissionDiagnosis
into #misscolonoscopy
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered is null
----(27399 row(s) affected)

select t1.*
from hed.dbo.HES_EPISODES t1 inner join #misscolonoscopy t2
                                                on t1.Patient_ID = t2.Patient_ID collate
database_default
                                                and t1.EPISTART = t2.firstAdmissionDiagnosis
order by t1.Patient_ID
-----16345
-----this looks interesting and confirms that paitnets seem to be having other kinds of scopes or
hemicolectmies etc
select t1.PAtient_ID, max(case when left(t1.opertn_01,3)  in ('H06', 'H07', 'H09' , 'H33', 'H20',
'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56', 'H11', 'H15', 'H21') then
case when opdate_01 IS not null then opdate_01 else EPISTART END
                                                        when left(t1.opertn_02,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_02 IS not null then opdate_02 else EPISTART END
                                                        when left(t1.opertn_03,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_03 IS not null then opdate_03 else EPISTART END
                                                        when left(t1.opertn_04,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_04 IS not null then opdate_04 else EPISTART END
                                                        when left(t1.opertn_05,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_05 IS not null then opdate_05 else EPISTART END
                                                        when left(t1.opertn_06,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_06 IS not null then opdate_06 else EPISTART END
                                                        when left(t1.opertn_07,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_07 IS not null then opdate_07 else EPISTART END
```

```
                                                    when left(t1.opertn_08,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_08 IS not null then opdate_08 else EPISTART END
                                                    when left(t1.opertn_09,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_09 IS not null then opdate_09 else EPISTART END
                                                    when left(t1.opertn_10,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_10 IS not null then opdate_10 else EPISTART END
                                                    when left(t1.opertn_11,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_11 IS not null then opdate_11 else EPISTART END
                                                    when left(t1.opertn_12,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_12 IS not null then opdate_12 else EPISTART END else null
end) colonoscopy_date
into  #updated_colorectaldate
from hed.dbo.HES_EPISODES t1 inner join #misscolonoscopy t2
                                                on t1.Patient_ID = t2.Patient_ID collate
database_default
                                                and t1.EPISTART = t2.firstAdmissionDiagnosis
group by t1.Patient_ID
---(15510 row(s) affected)
--and the other half of the data
select t1.PAtient_ID,max(case when left(t1.opertn_01,3)  in ('H06', 'H07', 'H09' , 'H33', 'H20',
'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56', 'H11', 'H15', 'H21') then
case when opdate_01 IS not null then opdate_01 else EPISTART END
                                                    when left(t1.opertn_02,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_02 IS not null then opdate_02 else EPISTART END
                                                    when left(t1.opertn_03,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_03 IS not null then opdate_03 else EPISTART END
                                                    when left(t1.opertn_04,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_04 IS not null then opdate_04 else EPISTART END
                                                    when left(t1.opertn_05,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_05 IS not null then opdate_05 else EPISTART END
                                                    when left(t1.opertn_06,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_06 IS not null then opdate_06 else EPISTART END
                                                    when left(t1.opertn_07,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_07 IS not null then opdate_07 else EPISTART END
                                                    when left(t1.opertn_08,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_08 IS not null then opdate_08 else EPISTART END
                                                    when left(t1.opertn_09,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_09 IS not null then opdate_09 else EPISTART END
                                                    when left(t1.opertn_10,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_10 IS not null then opdate_10 else EPISTART END
                                                    when left(t1.opertn_11,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_11 IS not null then opdate_11 else EPISTART END
                                                    when left(t1.opertn_12,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_12 IS not null then opdate_12 else EPISTART END else null
end) colonoscopy_date
into #updated_colorectaldate_P1
from hed.dbo.HES_EPISODES_Part1 t1 inner join #misscolonoscopy t2
                                                on t1.Patient_ID = t2.Patient_ID collate
database_default
                                                and t1.EPISTART = t2.firstAdmissionDiagnosis
group by t1.Patient_ID
---and then to update the data
---(9423 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = t2.colonoscopy_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join #updated_colorectaldate
t2
```

```
                                                                                on
t1.Patient_id = t2.patient_ID collate database_default
where Diag_date_altered is null
----(19617 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = t2.colonoscopy_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
#updated_colorectaldate_P1 t2
                                                                                on
t1.Patient_id = t2.patient_ID collate database_default
where Diag_date_altered is null
---(10484 row(s) affected)

---Lets have a look to see how many are left without adjusted dates
select distinct Patient_ID
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered is null
----10364

select distinct Patient_ID  , firstAdmissionDiagnosis
into #misscolonoscopy_2
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered is null
----10364
---Lets now look to see what is going on with these patietnds

select t1.Patient_ID, t1.ADMIDATE, t1.EPISTART, t2.firstAdmissionDiagnosis, t1.OPERTN_01,
t1.OPERTN_02, t1.OPERTN_03, t1.OPERTN_04, t1.OPERTN_05,
        t1.OPERTN_06, t1.OPERTN_07, t1.OPERTN_08, t1.OPERTN_09, t1.OPERTN_10, t1.OPERTN_11,
t1.opertn_12
from HED.dbo.HES_EPISODES t1 inner join #misscolonoscopy_2 t2
                                                        on t1.Patient_ID = t2.Patient_ID
                                                        and t1.EPISTART <= t2.firstAdmissionDiagnosis
order by t1.patient_ID, admidate

select t1.PAtient_ID, case when left(t1.opertn_01,3)  in ('H06', 'H07', 'H09' , 'H33', 'H20', 'H22',
'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56', 'H11', 'H15', 'H21') then  case
when opdate_01 IS not null then opdate_01 else EPISTART END
                                                        when left(t1.opertn_02,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_02 IS not null then opdate_02 else EPISTART END
                                                        when left(t1.opertn_03,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_03 IS not null then opdate_03 else EPISTART END
                                                        when left(t1.opertn_04,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_04 IS not null then opdate_04 else EPISTART END
                                                        when left(t1.opertn_05,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_05 IS not null then opdate_05 else EPISTART END
                                                        when left(t1.opertn_06,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_06 IS not null then opdate_06 else EPISTART END
                                                        when left(t1.opertn_07,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_07 IS not null then opdate_07 else EPISTART END
                                                        when left(t1.opertn_08,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_08 IS not null then opdate_08 else EPISTART END
                                                        when left(t1.opertn_09,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_09 IS not null then opdate_09 else EPISTART END
                                                        when left(t1.opertn_10,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_10 IS not null then opdate_10 else EPISTART END
                                                        when left(t1.opertn_11,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_11 IS not null then opdate_11 else EPISTART END
                                                        when left(t1.opertn_12,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_12 IS not null then opdate_12 else EPISTART END else null
end colonoscopy_date     ,
```

```
t1. OPERTN_01, t1.OPERTN_02, t1.OPERTN_03, t1.OPERTN_04, t1.OPERTN_05, t1.OPERTN_06, t1.OPERTN_07,
t1.OPERTN_08, t1.OPERTN_09, t1.opertn_10, t1.OPERTN_11, t1.opertn_12
into #updated_colorectaldate_P2
from hed.dbo.HES_EPISODEs t1 inner join #misscolonoscopy t2
                                        on t1.Patient_ID = t2.Patient_ID collate
database_default
                                        and datediff(dd, t1.EPISTART ,
t2.firstAdmissionDiagnosis) between 1 and 90
---there are a few updateable fields here
---     21651 row(s) affected)

select t1.PAtient_ID, case when left(t1.opertn_01,3)  in ('H06', 'H07', 'H09' , 'H33', 'H20', 'H22',
'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56', 'H11', 'H15', 'H21') then  case
when opdate_01 IS not null then opdate_01 else EPISTART END
                                                when left(t1.opertn_02,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_02 IS not null then opdate_02 else EPISTART END
                                                when left(t1.opertn_03,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_03 IS not null then opdate_03 else EPISTART END
                                                when left(t1.opertn_04,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_04 IS not null then opdate_04 else EPISTART END
                                                when left(t1.opertn_05,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_05 IS not null then opdate_05 else EPISTART END
                                                when left(t1.opertn_06,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_06 IS not null then opdate_06 else EPISTART END
                                                when left(t1.opertn_07,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_07 IS not null then opdate_07 else EPISTART END
                                                when left(t1.opertn_08,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_08 IS not null then opdate_08 else EPISTART END
                                                when left(t1.opertn_09,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_09 IS not null then opdate_09 else EPISTART END
                                                when left(t1.opertn_10,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_10 IS not null then opdate_10 else EPISTART END
                                                when left(t1.opertn_11,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_11 IS not null then opdate_11 else EPISTART END
                                                when left(t1.opertn_12,3)  in ('H06', 'H07',
'H09' , 'H33', 'H20',  'H22', 'H25', 'H28','H10', 'H23', 'H10', 'H41', 'H05', 'H24', 'H29', 'H56',
'H11', 'H15', 'H21') then case when opdate_12 IS not null then opdate_12 else EPISTART END else null
end colonoscopy_date     ,
t1. OPERTN_01, t1.OPERTN_02, t1.OPERTN_03, t1.OPERTN_04, t1.OPERTN_05, t1.OPERTN_06, t1.OPERTN_07,
t1.OPERTN_08, t1.OPERTN_09, t1.opertn_10, t1.OPERTN_11, t1.opertn_12
into #updated_colorectaldate_P3
from hed.dbo.HES_EPISODEs_part1 t1 inner join #misscolonoscopy t2
                                        on t1.Patient_ID = t2.Patient_ID collate
database_default
                                        and datediff(dd, t1.EPISTART ,
t2.firstAdmissionDiagnosis) between 1 and 90

---(16347 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = t2.colonoscopy_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
#updated_colorectaldate_P2 t2
                                                                        on
t1.Patient_id = t2.patient_ID collate database_default
where Diag_date_altered is null
----(5205 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = t2.colonoscopy_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
#updated_colorectaldate_P3 t2
                                                                        on
t1.Patient_id = t2.patient_ID collate database_default
where Diag_date_altered is null
```

```
---(3359 row(s) affected)

---lets see whether this has improved things
select distinct Patient_ID
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Diag_date_altered is null
----6138 rows - so we are definitely gewtting there
---I think the best thing to do for these patients is to set the diagnosis date to that of the first
admission date

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Diag_date_altered = firstAdmissionDiagnosis
where Diag_date_altered is null
---(7601 row(s) affected)

---Lets put in the altered tiem from colonoscopy
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add dayspostop_ColonoscpyDiag_altered int;

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set dayspostop_ColonoscpyDiag_altered = DATEDIFF(dd, colonoscopy_date, case when Diag_date_altered IS
null then firstAdmissionDiagnosis else Diag_date_altered end)
-----(84192 row(s) affected)

---I'll neeed to include an altered drop record now for the patients as I will no longer be excluding
patients who have their admission excluded as they are long-stayers
--Altough looking at hte data again I only excluded 8 patients so I think that I can live with them
being dropped

---Lets look to see whether we can link trusts
select distinct procode, case when procode = 'RY8' then 'Derbyshire Community Health Services NHS
Trust' else t2.[Trust name] end,
t3.[Trust Name], t4.procode3_mapped
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 left join
England.dbo.NHS_Trust_Details t2

                                on left(t1.Procode,3) = LEFT(t2.[trust code],3)
                                                                        --
                left join England.dbo.NHS_Trust_Details_additional t3

                                --on left(t1.Procode,3) = LEFT(t3.[code],3) -- this tabel only really
adds one thing to the records
                                                                                                left
join HED.dbo.ltbl_Organisation t3

                                on left(t1.Procode,3) = LEFT(t3.[Trust code],3)
                                                                                                left
join    [HED].[dbo].[ltbl_PROCODE3_MAPPED] t4

                                on t1.procode = t4.[procode3]
order by procode

---the mapped table does look like it could update a few of the records
alter table      Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add procode3_adjusted varchar(3);

 update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
 set procode3_adjusted = case when t2.procode3_adjusted Is not null then  t2.procode3_adjusted else
t1.procode end
 from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 left join
       [Cancer_research].[dbo].[ltblFE_20140109_ProcodeMapped] t2

                                on t1.procode = t2.procode
-----   (84192 row(s) affected)
---and lets now re-look at the provider names
select distinct procode3_adjusted, case when procode = 'RY8' then 'Derbyshire Community Health
Services NHS Trust' else t2.[Trust name] end,
t3.[Trust Name]
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 left join
England.dbo.NHS_Trust_Details t2

                                on left(t1.procode3_adjusted,3) = LEFT(t2.[trust code],3) collate
database_default
```

```
                                                                    --
                left join England.dbo.NHS_Trust_Details_additional t3

                                --on left(t1.Procode,3) = LEFT(t3.[code],3) -- this tabel only really
adds one thing to the records
                                                                                                    left
join HED.dbo.ltbl_Organisation t3

                                on left(t1.procode3_adjusted,3) = LEFT(t3.[Trust code],3)collate
database_default

order by procode3_adjusted
---as well as the PCTs there are an additional 12 providers which do not result in a trust name
---Lets look at Irena's table now

select *
from Irena.dbo.[Trust_Mod]
where LEFT([Provider_Code],3) in ('NT5', 'RCJ', 'REX', 'RG2', 'RGK', 'RH7', 'RHS', 'RMN', 'RMR',
'RR2', 'RR4', 'RVM')
---11 of these hqave been identified using Irena's table; But there are no new columns
--NT5 os careUK

select distinct LEFT(procode,3),spell_finYR
from HED.dbo.HES_SPELLS
where LEFT(procode,3) in ('NT5','NT6', 'RCJ', 'REX', 'RG2', 'RGK', 'RH7', 'RHS', 'RMN', 'RMR', 'RR2',
'RR4', 'RVM')
order by LEFT(procode,3),spell_finYR

---these all seem to be older providers NT5 is CareUK this may have to be left as is
select distinct Procode, procode3_adjusted
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses



select *
from HED.dbo.ltbl_Organisation t3
where t3.[Trust Name] like '%South West London%'

---WEs hould now be able to update 11/12 providers to a more recent one
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'R1F'
where procode = 'RR2';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RYQ'
where procode = 'RG2';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RTR'
where procode = 'RCJ';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RW6'
where procode in ('REX','RMN');
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RTQ'
where procode = 'RH7';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RT1'
where procode = 'RGK';
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'R1C'
where procode = 'RHS';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RXL'
where procode = 'RMR';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RXF'
where procode = 'RR4';
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'RQY'
where procode = 'RVM';
/*(53 row(s) affected)

(489 row(s) affected)

(0 row(s) affected)
```

```
(0 row(s) affected)

(0 row(s) affected)

(0 row(s) affected)

(0 row(s) affected)

(0 row(s) affected)

(0 row(s) affected)

(0 row(s) affected)

*/

---Lets have a look at how this looks now

select COUNT(*), procode3_adjusted, max(spell_finYR)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by procode3_adjusted
order by max(spell_finYR), procode3_adjusted
---Lets have a look at the trusts beginning with 'R' which have a final admisssion in 2001/02
select *
from HED.dbo.ltbl_Organisation a
where  a.[Trust Code] in ('R1C','RQY','RRE','RT1','RT5','RTQ','RWQ')
---these are all Mental health trusts, could this be something to do with patients who are
institutionalised, and therefore th eprovider looking after
---them would pay the coses, not necessarily the trust which physically provides the treatment
---there are some PCTs in the rest of the jumble lets look at the names for them
select distinct procode3_adjusted, t2.[Trust Name]
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 left join HED.dbo.ltbl_Organisation
t2
                                                                    on t1.procode3_adjusted = t2.[Trust
Code] collate database_default
order by procode3_adjusted

---there are 32 PCTS not all of which are linking to the provider table lets look at some of these
PCTS

select *
from Irena.dbo.Trust_Mod2
where Provider_code in ('5A1','5AJ','5C4','5CD','5CN','5CW','5DH','5DK',
                                                '5DQ','5F1','5FH','5FN','5FP','5FV','5FW','5JA',
                                                '5KR','5KY','5L9','5LG','5MT','5N6','5NV','5P6',
                                                '5PA','5PK','5PL','5QC','5QG','5QH','5QM','5QQ')
---They aren;t all in the original table, a few more are in the second table
select *
from England.dbo.PCT_LOOKUP_TABLE
where PCT_Code in ('5A1','5AJ','5C4','5CD','5CN','5CW','5DH','5DK',
                                                '5DQ','5F1','5FH','5FN','5FP','5FV','5FW','5JA',
                                                '5KR','5KY','5L9','5LG','5MT','5N6','5NV','5P6',
                                                '5PA','5PK','5PL','5QC','5QG','5QH','5QM','5QQ')
---even fewer of them in here...
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QC'
where procode in( '5A1', '5G6','5E9','5DF','5FD','5LX','5LY');
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5PV'
where procode = '5AJ';

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QM'
where procode in( '5CD', '5FN','5FP');
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QK'
where procode = '5DH';

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QF'
where procode = '5DK';
```

```
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5PK'
where procode = '5DQ';

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5P8'
where procode = '5FH';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QQ'
where procode = '5FV';

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QQ'
where procode = '5FV';

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QL'
where procode = '5FW';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5PA'
where procode = '5JA';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QP'
where procode = '5KR';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5QH'
where procode = '5KY';
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5P6'
where procode = '5L9';

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = '5PL'
where procode = '5MT'
/*(136 row(s) affected)

(1 row(s) affected)

(24 row(s) affected)

(0 row(s) affected)

(30 row(s) affected)

(28 row(s) affected)

(2 row(s) affected)

(5 row(s) affected)

(5 row(s) affected)

(0 row(s) affected)

(16 row(s) affected)

(1 row(s) affected)

(12 row(s) affected)

(1 row(s) affected)

(45 row(s) affected)

*/

---and again to check
select distinct procode3_adjusted, t2.[Trust Name]
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 left join HED.dbo.ltbl_Organisation
t2
                                                                    on t1.procode3_adjusted = t2.[Trust
Code] collate database_default
order by procode3_adjusted
---we are now only missing the one PCT and on private which is Care UK
```

```
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set procode3_adjusted = 'TAL'
where procode = '5CW'
---(1 row(s) affected)

select COUNT(*), procode3_adjusted, spell_finYR
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by procode3_adjusted, spell_finYR
order by  procode3_adjusted , spell_finYR

---The Royal Marsden is the only trust which is sticking out of the R ones  - apart from the metnal
health trusts we look ed at earlier
---After looking at the Royal  Marsden the first option is the flexible sigmoidoscopy
select procode3_adjusted ,min(spell_finYR)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by procode3_adjusted
order by  min(spell_finYR)
---RY8  RW1
select *
from  HED.dbo.ltbl_Organisation
where [Trust Code] in ('RY8',  'RW1')
/*SOUTHERN HEALTH NHS FOUNDATION TRUST - community health, mental health and learning disabilitiy
services
DERBYSHIRE COMMUNITY HEALTH SERVICES NHS TRUST - agaain this is a trust that provides day
case/outpatient services*/
---Need to investigate these two

set concat_null_yields_null off
select t1.Patient_ID, t1.droprecord
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_episodes_part1 t2
                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                and
t1.firstAdmissionDiagnosis = t2.epistart
where
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C181%' and side_cancer = 'Unknown'

---three records here which will need to be dropped
---three recordfs here
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in
('8922D0C8940332A641DC61F57CDEC426','4A2BDBA04531C18FF51070188BD97D35','946897F0C67CB5F8FC14A120B9579A
08')
----Lets have a look at the other HES episodes
select t1.Patient_ID, t1.droprecord
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2
                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                and
t1.firstAdmissionDiagnosis = t2.epistart
where
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C181%' and side_cancer = 'Right'


---556 further record her e
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in (select t1.Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2

                                            on t1.Patient_ID = t2.Patient_ID collate database_default

                                            and t1.firstAdmissionDiagnosis = t2.epistart
                                                where
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C181%' and side_cancer = 'Right'
```

```
                                                )
---(275 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in (select t1.Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes_part1
t2

                                        on t1.Patient_ID = t2.Patient_ID collate database_default

                                        and t1.firstAdmissionDiagnosis = t2.epistart
                                                where
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C181%' and side_cancer = 'Right'

                                                )
----(111 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in (select t1.Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes_part1
t2

                                        on t1.Patient_ID = t2.Patient_ID collate database_default

                                        and t1.firstAdmissionDiagnosis = t2.epistart
                                                where
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C181%' and side_cancer = 'Right'

                                                )
----(111 row(s) affected)
select t1.Patient_ID, t2.epistart,
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 diagnoses
into #colorectalcancer
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_episodes_part1 t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                        and
t1.firstAdmissionDiagnosis = t2.epistart
where
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' and side_cancer = 'Right'

---LEts now look at the spells table to see whether any of these cancers relate to the appendix
select max(case when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C181%'      then 1
                                when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C18[02345678]%' then 2              else 0 end ) ,
        t1.Patient_Id
from HED.dbo.HES_Spells t1 inner join    #colorectalcancer t2
                        on t1.Patient_ID = t2.Patient_ID collate database_default
                        and t1.admidate >= t2.EPISTART
group by t1.Patient_ID
having max(case when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C181%'      then 1
                                when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C18[02345678]%' then 2              else 0 end ) = 1
order by t1.patient_Id
----5 patients


---Now to look at further patients
select t1.Patient_ID, t2.epistart,
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 diagnoses
```

```
into #colorectalcancer_2
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                                    and
t1.firstAdmissionDiagnosis = t2.epistart
where
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' and side_cancer = 'Right'

-----(1307 row(s) affected)
---LEts now look at the spells table to see whether any of these cancers relate to the appendix
select max(case when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C181%'      then 1
                                 when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C18[02345678]%' then 2              else 0 end ) ,
        t1.Patient_Id
from HED.dbo.HES_Spells t1 inner join    #colorectalcancer_2 t2
                        on t1.Patient_ID = t2.Patient_ID collate database_default
                        and t1.admidate >= t2.EPISTART
group by t1.Patient_ID
having max(case when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C181%'      then 1
                                 when
t1.diag_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12 like '%C18[02345678]%' then 2              else 0 end ) = 1
order by t1.patient_Id
---4 patients here

----(5 row(s) affected)

---add cancer type

alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add type_cancer varchar(25);
---first investigsating side of the cancer i.e. left or right, Danny advised about whihch cancers are
considered either left or right
set concat_null_yields_null off
select t1.Patient_ID, case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
```

```
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        else null end
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_episodes_part1 t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                    and
t1.firstAdmissionDiagnosis = t2.epistart
                ---and to update

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer =   side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        else null end) side
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes_part1
t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                    and
t1.firstAdmissionDiagnosis = t2.epistart
                        group by t1.Patient_ID
        ) t2
                                                                on t1.Patient_ID = t2.Patient_ID

        -----    (40114 row(s) affected)
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
```

```
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        else null end) side
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                        and
t1.firstAdmissionDiagnosis = t2.epistart
                                        where   t1.type_cancer is null
                                        group by t1.PAtient_ID                                  ) t2
                                                        on t1.Patient_ID = t2.Patient_ID

----    (41106 row(s) affected)
---Let have a look at the results to check that everything update ok
select type_cancer, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by type_cancer
-----13301 nulls lets look at these
set concat_null_yields_null off
select
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 ,
case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
```

```
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        else null end side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_episodes_part1 t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                    and
t1.firstAdmissionDiagnosis = t2.epistart
where t1.droprecord is null and t1.type_cancer is null
----there are a lot which end in -9

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
```

```
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        else null end ) side
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                    and
t1.firstAdmissionDiagnosis < t2.epistart
                                    where droprecord is null and type_cancer is null
                                                            group by t1.patient_ID) t2
                                                    on t1.Patient_ID = t2.Patient_ID

        -----(8415 row(s) affected)

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        else null end) side
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes_part1
t2

                        on t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                    and
t1.firstAdmissionDiagnosis < t2.epistart
                                    where droprecord is null and type_cancer is null
                                                            group by t1.patient_ID) t2
                                                    on t1.Patient_ID = t2.Patient_ID
```

```
-----   (3898 row(s) affected)
---Lets check this out again
select type_cancer, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by type_cancer
---still 8.2k where unknown
---Lets  a  look to see if there are any records which have the cancer side on the death certificates

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C180%' then 'Caecum'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C182%' then 'Ascending colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C183%' then 'Hepatic Flexure'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C184%' then 'Transverse colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C185%' then 'Splenic Flexure'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C186%' then 'Descending colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C187%' then 'Sigmoid colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C188%' then 'Overlapping colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C19%' then 'Rectosigmoid Junction'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C20%' then 'Rectum'
```

```
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C210%' then 'Anus'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C211%' then 'Anal Canal'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C212%' then 'Cloacogenic zone' else null end ) side
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2

                        on t1.Patient_ID = t2.EXTRACT_HESID collate database_default
                                                                                and
t1.firstAdmissionDiagnosis < t2.DOD
                                    where droprecord is null and type_cancer is null
                                                            group by t1.patient_ID) t2
                                                    on t1.Patient_ID = t2.Patient_ID


----(3692 row(s) affected)
select type_cancer, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by type_cancer
---7869
set concat_null_yields_null off
select
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 ,
case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
```

```
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'
                        else null end side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_episodes_part1 t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID collate database_default
                                                                                    and
t1.firstAdmissionDiagnosis = t2.epistart
where t1.droprecord is null and t1.type_cancer is null

---these are almost all colrectal unknowns -
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'

                                                    when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D012%' then 'Carcinoma_rectum'

                                                    when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D010%' then 'Carcinoma_colon'

                                                                    else null end ) side

                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes_part1
t2


                        on t1.Patient_ID = t2.Patient_ID collate database_default

                                                                                    and
t1.firstAdmissionDiagnosis = t2.epistart

                                                where droprecord is null and type_cancer is null

                                                group by t1.patient_ID) t2
                                    on t1.Patient_ID = t2.Patient_ID
------  2474 row(s) affected

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'

                                                    when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D012%' then 'Carcinoma_rectum'

                                                    when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%D010%' then 'Carcinoma_colon'

                                                                    else null end ) side

                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HED.dbo.HES_episodes t2
```

```
                    on t1.Patient_ID = t2.Patient_ID collate database_default

                                                                                and
t1.firstAdmissionDiagnosis = t2.epistart

                                                where droprecord is null and type_cancer is null

                                                group by t1.patient_ID) t2
                                    on t1.Patient_ID = t2.Patient_ID
-----(2733 row(s) affected)

---Lets look to see whether we can find them in the mortality tables
select
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12
        ,case when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C180%' then 'Caecum'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C182%' then 'Ascending colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C183%' then 'Hepatic Flexure'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C184%' then 'Transverse colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C185%' then 'Splenic Flexure'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C186%' then 'Descending colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C187%' then 'Sigmoid colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C188%' then 'Overlapping colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C19%' then 'Rectosigmoid Junction'
```

```
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C20%' then 'Rectum'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C210%' then 'Anus'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C211%' then 'Anal Canal'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C189%' then 'Unspecified colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C212%' then 'Cloacogenic zone'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D100%' then 'Carcinoma_colon'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D101%' then 'Carcinoma_Rectosigmoid'
                        when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D102%' then 'Carcinoma_rectum' else null end
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                        on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                and t1.firstAdmissionDiagnosis = t2.DOD
where droprecord is null and type_cancer is null
---there are quite a few here...

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C180%' then 'Caecum'

                                                when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C182%' then 'Ascending colon'
```

```
                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C183%' then 'Hepatic Flexure'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C184%' then 'Transverse colon'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C185%' then 'Splenic Flexure'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C186%' then 'Descending colon'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C187%' then 'Sigmoid colon'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C188%' then 'Overlapping colon'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C19%' then 'Rectosigmoid Junction'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C20%' then 'Rectum'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C210%' then 'Anus'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C211%' then 'Anal Canal'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
```

```
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C189%' then 'Unspecified colon'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%C212%' then 'Cloacogenic zone'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D100%' then 'Carcinoma_colon'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D101%' then 'Carcinoma_Rectosigmoid'

                                                    when
t2.CAUSE_OF_DEATH_NON_NEONATAL_1+t2.CAUSE_OF_DEATH_NON_NEONATAL_2+t2.CAUSE_OF_DEATH_NON_NEONATAL_3+t2.
CAUSE_OF_DEATH_NON_NEONATAL_4+t2.CAUSE_OF_DEATH_NON_NEONATAL_5+t2.CAUSE_OF_DEATH_NON_NEONATAL_6+t2.CAU
SE_OF_DEATH_NON_NEONATAL_7+t2.CAUSE_OF_DEATH_NON_NEONATAL_8+t2.CAUSE_OF_DEATH_NON_NEONATAL_9+t2.CAUSE_
OF_DEATH_NON_NEONATAL_10+t2.CAUSE_OF_DEATH_NON_NEONATAL_11+t2.CAUSE_OF_DEATH_NON_NEONATAL_12 like
'%D102%' then 'Carcinoma_rectum' else null end ) side

                                from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1
inner join  HES_DW_ONS.dbo.HED_ONS_Mortality t2


                on t1.Patient_ID = t2.EXTRACT_HESID collate database_default

                                                                                    and
t1.firstAdmissionDiagnosis = t2.DOD

                                where droprecord is null and type_cancer is null

                                                    group by t1.patient_ID) t2
                                on t1.Patient_ID = t2.Patient_ID
-----(2433 row(s) affected)
---Lets have a look to see how many are remaingin
select type_cancer, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by type_cancer

----315


---now to try OP
set concat_null_yields_null off
select
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 ,
case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
```

```
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'
                        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'
                        else null end side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW.dbo.HES_OP_200607 t2
                                                                                on
t1.Patient_ID = t2.PatientID collate database_default
                                                                                and
t1.firstAdmissionDiagnosis = t2.apptdate
where t1.droprecord is null and t1.type_cancer is null
---there are some updateable fields here
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select PatientID ,

                                                                        case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'
```

```
        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'


        else null end side

                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HES_DW.dbo.HES_OP_200607 t2


                                                                on t1.Patient_ID =
t2.PatientID collate database_default


                                                                and
t1.firstAdmissionDiagnosis = t2.apptdate

                                                                        where t1.droprecord
is null and t1.type_cancer is null              ) t2

                        on t1.Patient_ID = t2.PatientID
------  (78 row(s) affected)


set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select PatientID ,

                                                                        case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'
```

```
        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'


        else null end side

                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HES_DW.dbo.HES_OP_200708 t2


                                                                on t1.Patient_ID =
t2.PatientID collate database_default


                                                                and
t1.firstAdmissionDiagnosis = t2.apptdate
```

```
                                                                                where t1.droprecord
is null and t1.type_cancer is null               ) t2

                         on t1.Patient_ID = t2.PatientID
-----(42 row(s) affected)


set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select PatientID ,

                                                                                case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'


        when
```

```
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'


        else null end side

                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HES_DW.dbo.HES_OP_200809 t2


                                                                        on t1.Patient_ID =
t2.PatientID collate database_default


                                                                        and
t1.firstAdmissionDiagnosis = t2.apptdate

                                                                                where t1.droprecord
is null and t1.type_cancer is null               ) t2

                        on t1.Patient_ID = t2.PatientID
-----    (27 row(s) affected)

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select PatientID ,

                                                                                case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'


        when
```

```
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'


        else null end side

                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HES_DW.dbo.HES_OP_200910 t2


                                                                on t1.Patient_ID =
t2.PatientID collate database_default


                                                                and
t1.firstAdmissionDiagnosis = t2.apptdate

                                                                        where t1.droprecord
is null and t1.type_cancer is null              ) t2

                        on t1.Patient_ID = t2.PatientID


----(35 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t2.Patient_ID
,

                                                                    case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'
```

```
        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'


        else null end side

                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HES_DW.dbo.HES_OP_201011 t2


                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default


                                                                    and
t1.firstAdmissionDiagnosis = t2.apptdate

                                                                        where t1.droprecord
is null and t1.type_cancer is null              ) t2

                        on t1.Patient_ID = t2.Patient_ID
----(33 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t2.Patient_ID
,
```

```
                                                                                case when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C180%' then 'Caecum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C182%' then 'Ascending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C183%' then 'Hepatic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C184%' then 'Transverse colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C185%' then 'Splenic Flexure'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C186%' then 'Descending colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C187%' then 'Sigmoid colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C188%' then 'Overlapping colon'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C19%' then 'Rectosigmoid Junction'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C20%' then 'Rectum'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C210%' then 'Anus'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C211%' then 'Anal Canal'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C212%' then 'Cloacogenic zone'


        when
t2.diag_01+t2.diag_02+t2.diag_03+t2.diag_04+t2.diag_05+t2.diag_06+t2.diag_07+t2.diag_08+t2.diag_09+t2.
diag_10+t2.diag_11+t2.diag_12 like '%C189%' then 'Unspecified colon'


        else null end side
```

```
                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  HES_DW.dbo.HES_OP_201112 t2

                                                                  on t1.Patient_ID =
t2.Patient_ID collate database_default

                                                                  and
t1.firstAdmissionDiagnosis = t2.apptdate

                                                                  where t1.droprecord
is null and t1.type_cancer is null               ) t2

                        on t1.Patient_ID = t2.Patient_ID
-----(24 row(s) affected)

select *
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer is null
---there look to be a few patients hwo have records mrore than once lets see how mnay paatients there
are in total and possibly carcinomas
select  distinct Patient_ID , firstAdmissionDiagnosis
into  #missingcolonoscopypatients
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer is null
---74
select t1.Patient_ID, t2.firstAdmissionDiagnosis, t1.ADMIDATE, t1.EPISTART, t1.DIAG_01, t1.DIAG_02,
t1.DIAG_03, t1.DIAG_04,
                t1.DIAG_05, t1.DIAG_06, t1.DIAG_07, t1.DIAG_08, t1.DIAG_09, t1.DIAG_10, t1.DIAG_11,
t1.DIAG_12 , t1.DIAG_13, t1.DIAG_14
                , t1.DIAG_15, t1.DIAG_16, t1.DIAG_17, t1.DIAG_18, t1.DIAG_19, t1.Diag_20, max(case
when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20 like '%C180%' then 'Caecum'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C182%' then 'Ascending colon'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C183%' then 'Hepatic Flexure'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C184%' then 'Transverse colon'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C185%' then 'Splenic Flexure'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C186%' then 'Descending colon'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C187%' then 'Sigmoid colon'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C188%' then 'Overlapping colon'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C19%' then 'Rectosigmoid Junction'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C20%' then 'Rectum'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C210%' then 'Anus'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C211%' then 'Anal Canal'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C212%' then 'Cloacogenic zone'
```

```
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D100%' then 'Carcinoma_colon'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D101%' then 'Carcinoma_Rectosigmoid'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D102%' then 'Carcinoma_rectum'
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C189%' then 'Unspecified colon'
                        else null end) side
from hed.dbo.HES_EPISODES t1 inner join  #missingcolonoscopypatients t2
                                    on t1.Patient_ID = t2.Patient_ID collate database_default
group by                t1.Patient_ID, t2.firstAdmissionDiagnosis, t1.ADMIDATE, t1.EPISTART,
t1.DIAG_01, t1.DIAG_02, t1.DIAG_03, t1.DIAG_04,
                t1.DIAG_05, t1.DIAG_06, t1.DIAG_07, t1.DIAG_08, t1.DIAG_09, t1.DIAG_10, t1.DIAG_11,
t1.DIAG_12 , t1.DIAG_13, t1.DIAG_14
                , t1.DIAG_15, t1.DIAG_16, t1.DIAG_17, t1.DIAG_18, t1.DIAG_19, t1.Diag_20
order by t1.patient_ID, t1.admidate
---there are a few who could be updated here
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20 like '%C180%' then 'Caecum'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C182%' then 'Ascending colon'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C183%' then 'Hepatic Flexure'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C184%' then 'Transverse colon'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C185%' then 'Splenic Flexure'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C186%' then 'Descending colon'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C187%' then 'Sigmoid colon'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C188%' then 'Overlapping colon'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C19%' then 'Rectosigmoid Junction'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C20%' then 'Rectum'
```

```
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C210%' then 'Anus'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C211%' then 'Anal Canal'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C212%' then 'Cloacogenic zone'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D100%' then 'Carcinoma_colon'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D101%' then 'Carcinoma_Rectosigmoid'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D102%' then 'Carcinoma_rectum'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C189%' then 'Unspecified colon'


                        else null end) side

                                                                                from
hed.dbo.HES_EPISODES t1 inner join  #missingcolonoscopypatients t2


                                on t1.Patient_ID = t2.Patient_ID collate database_default


        group by                t1.Patient_ID) t2

                                            on t1.Patient_ID = t2.Patient_ID

        -----    (95 row(s) affected)

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select t1.Patient_ID,
max(case when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20 like '%C180%' then 'Caecum'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C182%' then 'Ascending colon'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C183%' then 'Hepatic Flexure'


                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C184%' then 'Transverse colon'
```

```
                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C185%' then 'Splenic Flexure'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C186%' then 'Descending colon'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C187%' then 'Sigmoid colon'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C188%' then 'Overlapping colon'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C19%' then 'Rectosigmoid Junction'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C20%' then 'Rectum'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C210%' then 'Anus'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C211%' then 'Anal Canal'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C212%' then 'Cloacogenic zone'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D100%' then 'Carcinoma_colon'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D101%' then 'Carcinoma_Rectosigmoid'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%D102%' then 'Carcinoma_rectum'

                        when
diag_01+diag_02+diag_03+diag_04+diag_05+diag_06+diag_07+diag_08+diag_09+diag_10+diag_11+diag_12+diag_1
3+diag_14+diag_15+diag_16+diag_17+diag_18+diag_19+diag_20  like '%C189%' then 'Unspecified colon'

                        else null end) side

                                                                                from
hed.dbo.HES_EPISODES_part1 t1 inner join  #missingcolonoscopypatients t2

                                on t1.Patient_ID = t2.Patient_ID collate database_default
```

```
        group by                t1.Patient_ID) t2

                                                on t1.Patient_ID = t2.Patient_ID
where t1.type_cancer is null

------  (76 row(s) affected)
set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = side
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1   inner join (select
t1.EXTRACT_HESID,  max(case when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10 like '%C180%' then 'Caecum'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C182%' then 'Ascending colon'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C183%' then 'Hepatic Flexure'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C184%' then 'Transverse colon'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C185%' then 'Splenic Flexure'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C186%' then 'Descending colon'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C187%' then 'Sigmoid colon'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C188%' then 'Overlapping colon'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10  like '%C19%' then 'Rectosigmoid Junction'
```

```
                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C20%' then 'Rectum'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C210%' then 'Anus'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C211%' then 'Anal Canal'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C212%' then 'Cloacogenic zone'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%D100%' then 'Carcinoma_colon'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%D101%' then 'Carcinoma_Rectosigmoid'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10  like '%D102%' then 'Carcinoma_rectum'


                        when
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_
DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NE
ONATAL_10   like '%C189%' then 'Unspecified colon'


                        else null end) side

                                                                                from
HES_DW_ONS.dbo.HED_ONS_Mortality t1 inner join  #missingcolonoscopypatients t2

                                 on t1.EXTRACT_HESID = t2.Patient_ID collate database_default


        group by                t1.EXTRACT_HESID) t2

                                              on t1.Patient_ID = t2.EXTRACT_HESID      collate
database_default
where t1.type_cancer is null
-----(39 row(s) affected)
---We need to add the metasteses information now
```

```
select top 100 *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
---Lets add in the date of surgery  - I can then use this to define the mets issue

alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add date_surgery datetime,
LLung_mets int,
retroperti_perit_mets int,
Liver_mets int,
Bone_mets int ;
-----LEts update the date of suregery

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Date_surgery  = Operation_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID, MIN(
op_date)operation_date

                                        from HED.dbo.HES_spells t1 inner join
hed.dbo.HES_OPER_EXTRACT t2

                                                                                on
t1.spell_ID = t2.spell_ID

                                        where t2.OPER_CODE like 'H0[45][12389]%'

                                                or t2.oper_code like 'H0[67][123489]%'

                                                        or t2.oper_code like 'H0[89][1234589]%'

                                                        or t2.oper_code like 'H1[01][1234589]%'

                                                        or t2.oper_code like 'H12[2389]%'

                                                        or t2.oper_code like 'H29[123489]%'

                                                        or t2.oper_code like 'H33[123456789]%'

                                                        or t2.oper_code like 'H34[1589]%'

                                                        or t2.oper_code like 'H40[123]%'

                                    group by Patient_ID      ) t2
                                                                                on
t1.Patient_ID = t2.Patient_ID                                                   and
t1.firstAdmissionDiagnosis <= t2.OPeration_date
---(30454 row(s) affected)

SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Date_surgery  = Operation_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID, MIN(
op_date)operation_date

                                        from HED.dbo.HES_spells t1 inner join
hed.dbo.HES_OPER_EXTRACT_part1 t2

                                                                                on
t1.spell_ID = t2.spell_ID

                                        where t2.OPER_CODE like 'H0[45][12389]%'

                                                or t2.oper_code like 'H0[67][123489]%'

                                                        or t2.oper_code like 'H0[89][1234589]%'

                                                        or t2.oper_code like 'H1[01][1234589]%'

                                                        or t2.oper_code like 'H12[2389]%'

                                                        or t2.oper_code like 'H29[123489]%'
```

```
                                                    or t2.oper_code like 'H33[123456789]%'

                                                    or t2.oper_code like 'H34[1589]%'

                                                    or t2.oper_code like 'H40[123]%'

                                  group by Patient_ID      ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID                                                       and
t1.firstAdmissionDiagnosis <= t2.OPeration_date
---(27643 row(s) affected)
---Lets check the surgery date thing worked
select Patient_ID, surgery, Date_surgery
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Surgery = 1
Order by Date_surgery
---there are 367 patietns here who have had surgery but don't have a corresponding date for it
---Lets check there are no patients without surgery who have had a date of surgery


select Patient_ID, surgery, Date_surgery
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Surgery is null
Order by Date_surgery desc
---there aer  40 patients here - lets look at them (I suspect a lot of these are due to subsequent HES
updates)
select *
from HED.dbo.HES_spells
where Patient_ID in  ( select PAtient_ID
                                                  from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                  where surgery is null and date_surgery is not null
and droprecord is null )
order by patient_ID
---it looks like these people can have their records updated
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set surgery = 1
where Date_Surgery is not null and Surgery is null
---(44 row(s) affected)
---Lets have a look at tjhe records for one of hte patients who have a surgery record but no date

---this person has had one of the surgeries earlier than the date of diagnosis but has then
subsequently had a nother surgery
-----Lets alter the way this is updated
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Date_surgery  = Operation_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select t1.Patient_ID,
MIN(case when t3.op_date is not null then t3.op_date else t1.admidate end)        operation_date


                from HED.dbo.HES_SPELLS t1 inner join
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2

                                                    on t1.Patient_ID = t2.Patient_ID collate
database_default

                                                    and t1.admidate > =
t2.firstAdmissionDiagnosis

                                                                  inner join
HED.dbo.HES_OPER_EXtract t3

                                                            on t1.spell_ID = t2.spell_ID

                                        where t3.oper_code like '%H0[45][12389]%'

                                                or t3.oper_code like '%H0[67][123489]%'

                                                    or t3.oper_code like '%H0[89][1234589]%'

                                                    or t3.oper_code like '%H1[01][1234589]%'
```

```
                                                or t3.oper_code like '%H12[2389]%'
                                                or t3.oper_code like '%H29[123489]%'
                                                or t3.oper_code like '%H33[123456789]%'
                                                or t3.oper_code like '%H34[1589]%'
                                                or t3.oper_code like '%H40[123]%'
                                    group by t1.Patient_ID   ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.OPeration_date
Where surgery = 1 and date_surgery is null
---0 rows updated
SET CONCAT_NULL_YIELDS_NULL OFF
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Date_surgery  = Operation_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select t1.Patient_ID,
MIN(case when t3.op_date is not null then t3.op_date else t1.admidate end)        operation_date


                from HED.dbo.HES_SPELLS t1 inner join
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2

                                                on t1.Patient_ID = t2.Patient_ID collate
database_default

                                                and t1.admidate > =
t2.firstAdmissionDiagnosis

                                                                    inner join
HED.dbo.HES_OPER_EXtract_part1 t3

                                                            on t1.spell_ID = t2.spell_ID

                                        where t3.oper_code like '%H0[45][12389]%'

                                                or t3.oper_code like '%H0[67][123489]%'

                                                or t3.oper_code like '%H0[89][1234589]%'

                                                or t3.oper_code like '%H1[01][1234589]%'

                                                or t3.oper_code like '%H12[2389]%'

                                                or t3.oper_code like '%H29[123489]%'

                                                or t3.oper_code like '%H33[123456789]%'

                                                or t3.oper_code like '%H34[1589]%'

                                                or t3.oper_code like '%H40[123]%'

                                    group by t1.Patient_ID   ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.firstAdmissionDiagnosis <= t2.OPeration_date
Where surgery = 1 and date_surgery is null
---0 rows updated
-----(283 row(s) affected)
---Lets check to see how well this has improved things

select Patient_ID, surgery, Date_surgery
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Surgery = 1
Order by Date_surgery    , patient_ID
---25patietns  - although some of these are duplicates lets have a look at one of them

---there looks to have been a transciption error in the case of this operation
```

```
---Lets have a look to see if that is the case for the other patients
select Patient_ID, surgery, Date_surgery, firstAdmissionDiagnosis
into #colonsurgerymissdate
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Surgery = 1 and date_surgery is null
Order by Date_surgery    , patient_ID
-----(25 row(s) affected)
---Lets look at the records for these people
select t2.PAtient_ID, t2.SPELL_ID, t1.firstAdmissionDiagnosis, t2.OPERTN_01, t2.OPERTN_02,
t2.OPERTN_03, t2.OPERTN_03, t2.OPERTN_04, t2.OPERTN_05, t2.OPERTN_06, t2.OPERTN_07, t2.OPERTN_08,
        t2.OPERTN_09, t2.OPERTN_10, t2.OPERTN_11, t2.OPERTN_12, t2.opdate_01, t2.OPDATE_02,
t2.OPDATE_03, t2.OPDATE_04, t2.OPDATE_05, t2.EPISTART
from #colonsurgerymissdate t1 inner join HED.dbo.HES_EPISODES t2
                                                        on t1.Patient_ID = t2.Patient_ID collate
database_default
                                                        and t1.firstAdmissionDiagnosis <= t2.EPISTART
Order by Patient_ID, admidate
---there does seem to be transcription errors occurring here


---Lets see how this has done
select Patient_ID, surgery, Date_surgery
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Surgery = 1
Order by Date_surgery    , patient_ID


----this looks right now
----Can now think about adding in the mets info
set concat_null_yields_null off
select t1.Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
---16711*/
--Lets update these records
set concat_null_yields_null off
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
 -----(2622 row(s) affected)
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
----( 1571 row(s) affected)
---Lets look at adding in if the patient has primary lung cancer records
set concat_null_yields_null off
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C34%'
 -----(407 row(s) affected)
```

```
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C34%'
----(223 row(s) affected)


update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
 -----(1875 row(s) affected)
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
---(1216 row(s) affected)
---Now onto the liver I've seen some evidence about 25% of cancers have liver mets
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Liver_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and t1.firstAdmissionDiagnosis between t2.admidate and t2.disdate
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
 -----(2512 row(s) affected)
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Liver_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and t1.firstAdmissionDiagnosis between t2.admidate and t2.disdate
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
----(2333 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
 -----(791 row(s) affected)
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
```

```
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
----(405 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%c4[01]%'

---(4 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%c4[01]%'
----(10 row(s) affected)


update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%c48%'

----(27 row(s) affected)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%c48%'
----(67 row(s) affected)
---I also need to create a variable for looking at lymph node mets
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add lymphNodeMet int

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
 -----(5198 row(s) affected)
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstadmissionDiagnosis, t2.admidate) between 0
and 365
```

```
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
----(3975 row(s) affected)

---we haven't got anywhere near as amany patients with liver mets as I would expect - lets have a look
to see how many it increases to if we look at within the
--first year of diagnosis
alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add LiverMet_1YEar int
set concat_null_yields_null off
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstAdmissionDiagnosis , t2.ADMIDATE) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
 -----(6538 row(s) affected)
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstAdmissionDiagnosis , t2.ADMIDATE) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
----(5271 row(s) affected)

---check that there isn't any discrepancy between the two
select Liver_mets, LiverMet_1YEar, Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where Liver_mets  = 1 and LiverMet_1YEar is null
---about 484 patient
 update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar  = 1
where Liver_mets  = 1 and LiverMet_1YEar is null
----(484 row(s) affected)


select YEAR(firstAdmissionDiagnosis), COUNT(distinct Patient_ID), Count(distinct case when liver_mets
= 1 then Patient_ID End), COUNT(distinct case when LiverMet_1YEar = 1 then Patient_ID End)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by  YEAR(firstAdmissionDiagnosis)


---Lets have a look to see whterh anyone has also been diagnosed primary cancer of liver
select distinct t1.Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstAdmissionDiagnosis , t2.ADMIDATE) between 0
and 180
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C22%'
                        and LiverMet_1YEar is null
---this adds less than 23 patietns which isn't that many
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HED.dbo.HES_EPISODES_part1 t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstAdmissionDiagnosis , t2.ADMIDATE) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C22%'
-----(75 row(s) affected)
```

```
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HED.dbo.HES_EPISODES t2
                                on t1.Patient_ID = t2.Patient_ID collate database_default
                                and DATEDIFF(dd, t1.firstAdmissionDiagnosis , t2.ADMIDATE) between 0
and 365
where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C22%'
-----(128 row(s) affected)

---Lets have a look at records for people who died within the first year though
select distinct Patient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                        on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                                        and t1.dod = t2.DOD
where droprecord is null and LiverMet_1YEar is null and DATEDIFF(dd, FirstAdmissionDiagnosis, t1.DOD)
between 0 and 365 and (underlying in( 'C797', 'D015', 'D376') or cause_1 in( 'C797', 'D015', 'D376')
                or cause_2 in( 'C797', 'D015', 'D376') or cause_3 in( 'C797', 'D015', 'D376') or
CAUSE_OF_DEATH_NON_NEONATAL_4 in( 'C797', 'D015', 'D376')or CAUSE_OF_DEATH_NON_NEONATAL_5 in( 'C797',
'D015', 'D376')or CAUSE_OF_DEATH_NON_NEONATAL_6 in( 'C797', 'D015', 'D376')
                or CAUSE_OF_DEATH_NON_NEONATAL_7 in( 'C797', 'D015', 'D376')or
CAUSE_OF_DEATH_NON_NEONATAL_8 in( 'C797', 'D015', 'D376')or CAUSE_OF_DEATH_NON_NEONATAL_9 in( 'C797',
'D015', 'D376') or CAUSE_OF_DEATH_NON_NEONATAL_10 in( 'C797', 'D015', 'D376')
                or CAUSE_OF_DEATH_NON_NEONATAL_11 in( 'C797', 'D015', 'D376') or
CAUSE_OF_DEATH_NON_NEONATAL_12 in( 'C797', 'D015', 'D376')
                or      underlying like'C22%' or cause_1 like 'C22%'     or cause_2 like 'C22%' or
cause_3  like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_4 like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_5
like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_6  like 'C22%'
                or CAUSE_OF_DEATH_NON_NEONATAL_7  like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_8  like
'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_9  like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_10  like 'C22%'
                or CAUSE_OF_DEATH_NON_NEONATAL_11  like 'C22%' or CAUSE_OF_DEATH_NON_NEONATAL_12  like
'C22%')
-----13 rows
--(5 row(s) affected)
---and for the other types of mets
select distinct Patient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                        on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                                        and t1.dod = t2.DOD
where droprecord is null and lymphNodeMet is null and DATEDIFF(dd, FirstAdmissionDiagnosis, t1.DOD)
between 0 and 365 and (underlying like '%C77[1245]%' or cause_1  like '%C77[1245]%'
                or cause_2  like '%C77[1245]%' or cause_3  like '%C77[1245]%' or
CAUSE_OF_DEATH_NON_NEONATAL_4  like '%C77[1245]%' or CAUSE_OF_DEATH_NON_NEONATAL_5 like
'%C77[1245]%'or CAUSE_OF_DEATH_NON_NEONATAL_6  like '%C77[1245]%'
                or CAUSE_OF_DEATH_NON_NEONATAL_7  like '%C77[1245]%'or CAUSE_OF_DEATH_NON_NEONATAL_8
like '%C77[1245]%' or CAUSE_OF_DEATH_NON_NEONATAL_9  like '%C77[1245]%'or
CAUSE_OF_DEATH_NON_NEONATAL_10  like '%C77[1245]%'
                or CAUSE_OF_DEATH_NON_NEONATAL_11  like '%C77[1245]%' or
CAUSE_OF_DEATH_NON_NEONATAL_12  like '%C77[1245]%'
                )
----5 rows

select distinct Patient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                        on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                                        and t1.dod = t2.DOD
where droprecord is null and Bone_mets is null and DATEDIFF(dd, FirstAdmissionDiagnosis, t1.DOD)
between 0 and 365 and (underlying like '%C795%' or cause_1  like '%C795%'
                or cause_2  like '%C795%' or cause_3  like '%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_4
like '%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_5 like '%C795%'or CAUSE_OF_DEATH_NON_NEONATAL_6  like
'%C795%'
                or CAUSE_OF_DEATH_NON_NEONATAL_7  like '%C795%'or CAUSE_OF_DEATH_NON_NEONATAL_8  like
'%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_9  like '%C795%'or CAUSE_OF_DEATH_NON_NEONATAL_10  like
'%C795%'
                or CAUSE_OF_DEATH_NON_NEONATAL_11  like '%C795%' or CAUSE_OF_DEATH_NON_NEONATAL_12
like '%C795%'
                )
```

```
                ---30 rows

select distinct Patient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                    on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                                    and t1.dod = t2.DOD
where droprecord is null and retroperti_perit_mets is null and DATEDIFF(dd, FirstAdmissionDiagnosis,
t1.DOD) between 0 and 365 and (underlying like '%C786%' or cause_1  like '%C786%'
                or cause_2  like '%C786%' or cause_3  like '%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_4
like '%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_5 like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_6  like
'%C786%'
                or CAUSE_OF_DEATH_NON_NEONATAL_7  like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_8  like
'%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_9  like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_10  like
'%C786%'
                or CAUSE_OF_DEATH_NON_NEONATAL_11  like '%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_12
like '%C786%'
                )
---158
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID  in (select distinct Patient_ID
                                                        from
        Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                    on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                                    and t1.dod = t2.DOD
                                                    where droprecord is null and
retroperti_perit_mets is null and DATEDIFF(dd, FirstAdmissionDiagnosis, t1.DOD) between 0 and 365 and
(underlying like '%C786%' or cause_1  like '%C786%'
                                                                    or cause_2  like
'%C786%' or cause_3  like '%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_4  like '%C786%' or
CAUSE_OF_DEATH_NON_NEONATAL_5 like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_6  like '%C786%'
                                                                    or
CAUSE_OF_DEATH_NON_NEONATAL_7  like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_8  like '%C786%' or
CAUSE_OF_DEATH_NON_NEONATAL_9  like '%C786%'or CAUSE_OF_DEATH_NON_NEONATAL_10  like '%C786%'
                                                                    or
CAUSE_OF_DEATH_NON_NEONATAL_11  like '%C786%' or CAUSE_OF_DEATH_NON_NEONATAL_12  like '%C786%'
                                                                    ))
----(55 row(s) affected)
select distinct Patient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                    on t1.Patient_ID =
t2.EXTRACT_HESID collate database_default
                                                                    and t1.dod = t2.DOD
where droprecord is null and LLung_mets is null and DATEDIFF(dd, FirstAdmissionDiagnosis, t1.DOD)
between 0 and 365 and (underlying in( 'C780', 'C34') or cause_1 in('C780', 'C34' )
                or cause_2 in( 'C780', 'C34') or cause_3 in('C780', 'C34') or
CAUSE_OF_DEATH_NON_NEONATAL_4 in( 'C780', 'C34')or CAUSE_OF_DEATH_NON_NEONATAL_5 in('C780', 'C34')or
CAUSE_OF_DEATH_NON_NEONATAL_6 in('C780', 'C34')
                or CAUSE_OF_DEATH_NON_NEONATAL_7 in( 'C780', 'C34')or CAUSE_OF_DEATH_NON_NEONATAL_8
in( 'C780', 'C34')or CAUSE_OF_DEATH_NON_NEONATAL_9 in('C780', 'C34') or CAUSE_OF_DEATH_NON_NEONATAL_10
in('C780', 'C34')
                or CAUSE_OF_DEATH_NON_NEONATAL_11 in( 'C780', 'C34') or CAUSE_OF_DEATH_NON_NEONATAL_12
in( 'C780', 'C34')
                )

----227
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
where Patient_ID  in (select distinct Patient_ID
                                        from
        Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2

                        on t1.Patient_ID = t2.EXTRACT_HESID collate database_default

                        and t1.dod = t2.DOD
                                        where droprecord is null and LLung_mets is null and
DATEDIFF(dd, FirstAdmissionDiagnosis, t1.DOD) between 0 and 365 and (underlying in( 'C780', 'C34') or
cause_1 in('C780', 'C34' )
```

```
                                                    or cause_2 in( 'C780', 'C34') or cause_3
in('C780', 'C34') or CAUSE_OF_DEATH_NON_NEONATAL_4 in( 'C780', 'C34')or CAUSE_OF_DEATH_NON_NEONATAL_5
in('C780', 'C34')or CAUSE_OF_DEATH_NON_NEONATAL_6 in('C780', 'C34')
                                                    or CAUSE_OF_DEATH_NON_NEONATAL_7 in( 'C780',
'C34')or CAUSE_OF_DEATH_NON_NEONATAL_8 in( 'C780', 'C34')or CAUSE_OF_DEATH_NON_NEONATAL_9 in('C780',
'C34') or CAUSE_OF_DEATH_NON_NEONATAL_10 in('C780', 'C34')
                                                    or CAUSE_OF_DEATH_NON_NEONATAL_11 in( 'C780',
'C34') or CAUSE_OF_DEATH_NON_NEONATAL_12 in( 'C780', 'C34')
                                                    ))
----(152 row(s) affected)


update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200607 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
                                                    and LiverMet_1YEar is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200708 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
                                                    and LiverMet_1YEar is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200809 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
                                                    and LiverMet_1YEar is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200910 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
                                                    and LiverMet_1YEar is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201011 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
```

```
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
                                                        and LiverMet_1YEar is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201112 t2
                                                                on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
                                                        and LiverMet_1YEar is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LiverMet_1YEar = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201213 t2
                                                                on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C787%'
                                                        and LiverMet_1YEar is null)



update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200607 t2
                                                                on t1.Patient_ID =
t2.PatientID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
                                                        and Bone_mets is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200708 t2
                                                                on t1.Patient_ID =
t2.PatientID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
                                                        and Bone_mets is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200809 t2
                                                                on t1.Patient_ID =
t2.PatientID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
```

```
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
                                                        and Bone_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200910 t2
                                                                on t1.Patient_ID =
t2.PatientID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
                                                        and Bone_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201011 t2
                                                                on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
                                                        and Bone_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201112 t2
                                                                on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
                                                        and Bone_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Bone_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201213 t2
                                                                on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C795%'
                                                        and Bone_mets is null)


update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200607 t2
                                                                on t1.Patient_ID =
t2.PatientID collate database_default
                                                                and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
                                                        and retroperti_perit_mets is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
```

```
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200708 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
                                                        and retroperti_perit_mets is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200809 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
                                                        and retroperti_perit_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200910 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
                                                        and retroperti_perit_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201011 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
                                                        and retroperti_perit_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201112 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
                                                        and retroperti_perit_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set retroperti_perit_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201213 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
```

```
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C786%'
                                                    and retroperti_perit_mets is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200607 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
                                                    and LLung_mets is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200708 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
                                                    and LLung_mets is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200809 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
                                                    and LLung_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200910 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
                                                    and LLung_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201011 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
                                                    and LLung_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
```

```
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201112 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
                                                        and LLung_mets is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LLung_mets = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201213 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C780%'
                                                        and LLung_mets is null)



update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200607 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
                                                        and lymphNodeMet is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200708 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
                                                        and lymphNodeMet is null)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200809 t2
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                    where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
                                                        and lymphNodeMet is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID in (select distinct t1.Patient_ID
                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_200910 t2
```

```
                                                                    on t1.Patient_ID =
t2.PatientID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
                                                    and lymphNodeMet is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201011 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
                                                    and lymphNodeMet is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201112 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
                                                    and lymphNodeMet is null)
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lymphNodeMet = 1
where Patient_ID in (select distinct t1.Patient_ID
                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join HES_DW.dbo.HES_OP_201213 t2
                                                                    on t1.Patient_ID =
t2.Patient_ID collate database_default
                                                                    and DATEDIFF(dd,
t1.firstAdmissionDiagnosis , t2.APPTDATE) between 0 and 365
                                        where
t2.DIAG_01+t2.DIAG_02+t2.DIAG_03+t2.DIAG_04+t2.DIAG_05+t2.DIAG_06+t2.DIAG_07+t2.DIAG_08+t2.DIAG_09+t2.
DIAG_10+t2.DIAG_11+t2.DIAG_12 like '%C77[1245]%'
                                                    and lymphNodeMet is null)



---Lets go back to the type of cancer


alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add day_colonoscopy varchar(10);



update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set day_colonoscopy = DATENAME(dw, colonoscopy_date)
---(160924 row(s) affected)
----Lets go bqck to the cancer sites and double check these records
select *
from #missingcolonoscopypatients t1 inner join HED.dbo.HES_EPISODES t2
                                        on t1.Patient_ID = t2.PAtient_Id collate database_default
                                        and t1.firstAdmissionDiagnosis = t2.Epistart
-----   only  patients here, both with malignant neoplasm of appendix - so we need to drop these...

update   Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where Patient_ID in ()
```

```
select *
from #missingcolonoscopypatients t1 inner join HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                        on t1.Patient_ID = t2.EXTRACT_HESID collate database_default
                                        and t1.firstAdmissionDiagnosis = t2.DOD
------2 patients here have carcinoma the rest have appendix
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = 'Carcinoma_colon'
where PAtient_ID in ()
-

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where PAtient_ID in (select t1.Patient_ID
                                                from #missingcolonoscopypatients t1 inner join
HES_DW_ONS.dbo.HED_ONS_Mortality t2
                                                                                        on
t1.Patient_ID = t2.EXTRACT_HESID collate database_default
                                                                                        and
t1.firstAdmissionDiagnosis = t2.DOD
                                        where Patient_ID not in
----(29 row(s) affected)

----lets redo the tavle and see where we are at

drop table #missingcolonoscopypatients
select  distinct Patient_ID , firstAdmissionDiagnosis
into  #missingcolonoscopypatients
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer is null
----    (62 row(s) affected)
select *
from #missingcolonoscopypatients t1 inner join HES_DW.dbo.HES_OP_200607 t2
                                        on t1.Patient_ID = t2.PatientID collate database_default
                                        and t1.firstAdmissionDiagnosis = t2.Apptdate
---three patients to uypdate here


update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set droprecord = 1
where type_cancer in( 'Anus', 'Anal Canal')
----(47 row(s) affected)

select  type_cancer, side_cancer, COUNT(distinct Patient_ID)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by type_cancer, side_cancer
Order by  type_cancer, side_cancer

---some of these sides have gotten a bit muddled

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = 'Left'
where type_cancer in ('Carcinoma_Rectosigmoid','Carcinoma_rectum','Descending colon','Rectosigmoid
Junction','Rectum','Sigmoid colon','Splenic Flexure')
----(91943 row(s) affected)

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = 'Right'
where type_cancer in ('Ascending colon','Caecum','Hepatic Flexure','Transverse colon')
----(55308 row(s) affected)
---Lets have a look at the people who are in the unspecified colon group

select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
---9473
set concat_null_yields_null off
select t1.*
into  #additionalColonsCancerCodes
from HES_DW.dbo.HES_OP_200607  t1 inner join #UnspefColonPatients t2
                                        on t1.PATIENTID = t2.Patient_ID collate
database_default
                                        and t1.APPTDATE > t2.firstAdmissionDiagnosis
```

```
where DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
like '%c18[02345678]%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c19%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c20%'
Order by t1.PATIENTID, t1.APPTDATE
---                     (73 row(s) affected)
select distinct  PatientID, case when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C180%' then 'C180'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C182%' then 'C182'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C183%' then 'C183'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C184%' then 'C184'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C185%' then 'C185'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C186%' then 'C186'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C187%' then 'C187'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C188%' then 'C188'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C19%' then 'C19'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'%C20%' then 'C20'
                                                when
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12 like
'C180' then 'C180' end
from      #additionalColonsCancerCodes
---equates to nine rows


drop table #UnspefColonPatients



select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
---9464
drop table  #additionalColonsCancerCodes

set concat_null_yields_null off
select t1.PatientID,
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
Diagnoses
into  #additionalColonsCancerCodes
from HES_DW.dbo.HES_OP_200708  t1 inner join #UnspefColonPatients t2
                                                on t1.PATIENTID = t2.Patient_ID collate
database_default
                                                and t1.APPTDATE > t2.firstAdmissionDiagnosis
where DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
like '%c18[02345678]%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c19%' or
```

```
        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c20%'
Order by t1.PATIENTID, t1.APPTDATE
----(73 row(s) affected)
select distinct PatientID,case when Diagnoses like '%C180%' then 'C180'
                                                when Diagnoses like '%C182%' then 'C182'
                                                when Diagnoses like '%C183%' then 'C183'
                                                when Diagnoses like '%C184%' then 'C184'
                                                when Diagnoses like '%C185%' then 'C185'
                                                when Diagnoses like '%C186%' then 'C186'
                                                when Diagnoses like '%C187%' then 'C187'
                                                when Diagnoses like '%C188%' then 'C188'
                                                when Diagnoses like '%C19%' then 'C19'
                                                when Diagnoses like '%C20%' then 'C20' end
from     #additionalColonsCancerCodes
---7 rows


drop table #UnspefColonPatients
select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
---9457
drop table  #additionalColonsCancerCodes
set concat_null_yields_null off
select t1.PatientID,
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
Diagnoses
into  #additionalColonsCancerCodes
from HES_DW.dbo.HES_OP_200809  t1 inner join #UnspefColonPatients t2
                                                on t1.PATIENTID = t2.Patient_ID collate
database_default
                                                and t1.APPTDATE > t2.firstAdmissionDiagnosis
where DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
like '%c18[02345678]%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c19%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c20%'
Order by t1.PATIENTID, t1.APPTDATE
----(160 row(s) affected)
select distinct PatientID,case when Diagnoses like '%C180%' then 'C180'
                                                when Diagnoses like '%C182%' then 'C182'
                                                when Diagnoses like '%C183%' then 'C183'
                                                when Diagnoses like '%C184%' then 'C184'
                                                when Diagnoses like '%C185%' then 'C185'
                                                when Diagnoses like '%C186%' then 'C186'
                                                when Diagnoses like '%C187%' then 'C187'
                                                when Diagnoses like '%C188%' then 'C188'
                                                when Diagnoses like '%C19%' then 'C19'
                                                when Diagnoses like '%C20%' then 'C20' end
from     #additionalColonsCancerCodes
----14 rows

drop table #UnspefColonPatients
select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
drop table  #additionalColonsCancerCodes
set concat_null_yields_null off
select t1.PatientID,
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
Diagnoses
into  #additionalColonsCancerCodes
from HES_DW.dbo.HES_OP_200910  t1 inner join #UnspefColonPatients t2
                                                on t1.PATIENTID = t2.Patient_ID collate
database_default
                                                and t1.APPTDATE > t2.firstAdmissionDiagnosis
```

```
where DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
like '%c18[02345678]%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c19%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c20%'
Order by t1.PATIENTID, t1.APPTDATE
----(46 row(s) affected)
select distinct PatientID,case when Diagnoses like '%C180%' then 'C180'
                                                when Diagnoses like '%C182%' then 'C182'
                                                when Diagnoses like '%C183%' then 'C183'
                                                when Diagnoses like '%C184%' then 'C184'
                                                when Diagnoses like '%C185%' then 'C185'
                                                when Diagnoses like '%C186%' then 'C186'
                                                when Diagnoses like '%C187%' then 'C187'
                                                when Diagnoses like '%C188%' then 'C188'
                                                when Diagnoses like '%C19%' then 'C19'
                                                when Diagnoses like '%C20%' then 'C20' end
from     #additionalColonsCancerCodes


drop table #UnspefColonPatients
select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
---9436
drop table  #additionalColonsCancerCodes
set concat_null_yields_null off
select t1.Patient_ID,
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
Diagnoses
into  #additionalColonsCancerCodes
from HES_DW.dbo.HES_OP_201112  t1 inner join #UnspefColonPatients t2
                                                on t1.PATIENT_ID = t2.Patient_ID collate
database_default
                                                and t1.APPTDATE > t2.firstAdmissionDiagnosis
where DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
like '%c18[02345678]%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c19%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c20%'
Order by t1.PATIENT_ID, t1.APPTDATE
----(79 row(s) affected)
select distinct Patient_ID,case when Diagnoses like '%C180%' then 'C180'
                                                when Diagnoses like '%C182%' then 'C182'
                                                when Diagnoses like '%C183%' then 'C183'
                                                when Diagnoses like '%C184%' then 'C184'
                                                when Diagnoses like '%C185%' then 'C185'
                                                when Diagnoses like '%C186%' then 'C186'
                                                when Diagnoses like '%C187%' then 'C187'
                                                when Diagnoses like '%C188%' then 'C188'
                                                when Diagnoses like '%C19%' then 'C19'
                                                when Diagnoses like '%C20%' then 'C20' end
from     #additionalColonsCancerCodes
----7
drop table #UnspefColonPatients
select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
---9429

drop table  #additionalColonsCancerCodes
set concat_null_yields_null off
select t1.Patient_ID,
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
Diagnoses
into  #additionalColonsCancerCodes
```

```
from HES_DW.dbo.HES_OP_201213  t1 inner join #UnspefColonPatients t2
                                                on t1.PATIENT_ID = t2.Patient_ID collate
database_default
                                                and t1.APPTDATE > t2.firstAdmissionDiagnosis
where DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
like '%c18[02345678]%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c19%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c20%'
Order by t1.PATIENT_ID, t1.APPTDATE
----(44 row(s) affected)
select distinct Patient_ID,case when Diagnoses like '%C180%' then 'C180'
                                                when Diagnoses like '%C182%' then 'C182'
                                                when Diagnoses like '%C183%' then 'C183'
                                                when Diagnoses like '%C184%' then 'C184'
                                                when Diagnoses like '%C185%' then 'C185'
                                                when Diagnoses like '%C186%' then 'C186'
                                                when Diagnoses like '%C187%' then 'C187'
                                                when Diagnoses like '%C188%' then 'C188'
                                                when Diagnoses like '%C19%' then 'C19'
                                                when Diagnoses like '%C20%' then 'C20' end
from     #additionalColonsCancerCodes



drop table #UnspefColonPatients
select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
---9423

drop table  #additionalColonsCancerCodes
set concat_null_yields_null off
select t1.Patient_ID,
DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
Diagnoses
into  #additionalColonsCancerCodes
from HES_DW.dbo.HES_OP_201314  t1 inner join #UnspefColonPatients t2
                                                on t1.PATIENT_ID = t2.Patient_ID collate
database_default
                                                and t1.APPTDATE > t2.firstAdmissionDiagnosis
where DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_12
like '%c18[02345678]%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c19%' or

        DIAG_01+DIAG_02+DIAG_03+DIAG_04+DIAG_05+diag_06+diag_07+DIAG_08+DIAG_09+DIAG_10+DIAG_11+DIAG_1
2 like '%c20%'
Order by t1.PATIENT_ID, t1.APPTDATE
----(22 row(s) affected)
select distinct Patient_ID,case when Diagnoses like '%C180%' then 'C180'
                                                when Diagnoses like '%C182%' then 'C182'
                                                when Diagnoses like '%C183%' then 'C183'
                                                when Diagnoses like '%C184%' then 'C184'
                                                when Diagnoses like '%C185%' then 'C185'
                                                when Diagnoses like '%C186%' then 'C186'
                                                when Diagnoses like '%C187%' then 'C187'
                                                when Diagnoses like '%C188%' then 'C188'
                                                when Diagnoses like '%C19%' then 'C19'
                                                when Diagnoses like '%C20%' then 'C20' end
from     #additionalColonsCancerCodes


drop table #UnspefColonPatients
select distinct patient_ID, firstAdmissionDiagnosis
into #UnspefColonPatients
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer = 'Unspecified colon'
---9422
```

```
drop table  #additionalColonsCancerCodes
set concat_null_yields_null off
select t2.Patient_ID,
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5
+CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAUSE_OF_DE
ATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON_NEONATAL_11
+CAUSE_OF_DEATH_NON_NEONATAL_12 Diagnoses
into  #additionalColonsCancerCodes
from HES_DW_ONS.dbo.HED_ONS_Mortality  t1 inner join #UnspefColonPatients t2
                                                on t1.EXTRACT_HESID = t2.Patient_ID collate
database_default
                                                and t1.dod > t2.firstAdmissionDiagnosis
where
CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_NEONATAL
_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5

        +CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAU
SE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON_NEONATAL_11

        +CAUSE_OF_DEATH_NON_NEONATAL_12 like '%c18[02345678]%' or

        CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_
NEONATAL_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5

        +CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAU
SE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON_NEONATAL_11
                                        +CAUSE_OF_DEATH_NON_NEONATAL_12 like '%c19%' or

        CAUSE_OF_DEATH+CAUSE_OF_DEATH_NON_NEONATAL_1+CAUSE_OF_DEATH_NON_NEONATAL_2+CAUSE_OF_DEATH_NON_
NEONATAL_3+CAUSE_OF_DEATH_NON_NEONATAL_4+CAUSE_OF_DEATH_NON_NEONATAL_5

        +CAUSE_OF_DEATH_NON_NEONATAL_6+CAUSE_OF_DEATH_NON_NEONATAL_7+CAUSE_OF_DEATH_NON_NEONATAL_8+CAU
SE_OF_DEATH_NON_NEONATAL_9+CAUSE_OF_DEATH_NON_NEONATAL_10+CAUSE_OF_DEATH_NON_NEONATAL_11
                                    +CAUSE_OF_DEATH_NON_NEONATAL_12 like '%c20%'

Order by t2.PATIENT_ID, t1.DOD
----(0 row(s) affected)
---Lets now update the sides again
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = 'Left'
where type_cancer in ('Carcinoma_Rectosigmoid','Carcinoma_rectum','Descending colon','Rectosigmoid
Junction','Rectum','Sigmoid colon','Splenic Flexure')
----(91970 row(s) affected)

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = 'Right'
where type_cancer in ('Ascending colon','Caecum','Hepatic Flexure','Transverse colon')
----(55339 row(s) affected)

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set side_cancer = 'Unknown'
where type_cancer in ('Unspecified colon','Carcinoma_colon','Overlapping colon')
----(13651 row(s) affected)
select distinct type_Cancer, side_cancer, count(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by type_Cancer, side_cancer
order by type_cancer
---there is one person here who doesn't have a type pf cancer

select Patient_ID, firstAdmissionDiagnosis
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and type_cancer is null
----2009-09-11 00:00:00.000

---This data has led to a reduction in the probably misseds
---LEts compare the two tabel to see where the  differences are

select Patient_ID, firstAdmissionDiagnosis, colonoscopy_date
into #additionalPatientsfromEarliertable
From Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
```

```
where droprecord is null and LastColonoscopybeforeDiag =1  and IBD = 0 and
[dayspostopColonoscopyDiag]between 366 and 1096
-----    (8195 row(s) affected)

select t1.Patient_ID, t1.droprecord, t1.dayspostopColonoscopyDiag, t2.firstAdmissionDiagnosis,
t1.firstAdmissionDiagnosis , t1.LastColonoscopybeforeDiag
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
#additionalPatientsfromEarliertable t2
                                                        on t1.Patient_ID = t2.Patient_ID collate
database_default
                                                        and t1.colonoscopy_date = t2.colonoscopy_date
---there are a few records here which have not been coded as last colonoscopy before diag , but
totherwise I'm not sure why they aren't in this group
select t1.Patient_ID, t1.droprecord, t1.dayspostopColonoscopyDiag, t2.firstAdmissionDiagnosis,
t1.LastColonoscopybeforeDiag
into #QueryLastColonscopy
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
#additionalPatientsfromEarliertable t2
                                                        on t1.Patient_ID = t2.Patient_ID collate
database_default
                                                        and t1.colonoscopy_date = t2.colonoscopy_date
where   LastColonoscopybeforeDiag = 0
----(2498 row(s) affected)

select distinct t3.LastColonoscopybeforeDiag, t3.firstAdmissionDiagnosis, t1.*
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join #QueryLastColonscopy t2
                                        on t1.Patient_ID = t2.Patient_ID collate database_default

                                        left join #QueryLastColonscopy t3
                                        on t1.Patient_ID = t3.Patient_ID collate database_default

                                        and t1.dayspostopColonoscopyDiag =
t3.dayspostopColonoscopyDiag
Order by t1.PAtient_ID, t1.colonoscopy_date

---I've realised what the problem is - we redefined the "last" colonsocopy so I need to rerun these
analyses

select Patient_ID, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by patient_ID
Order by COUNT(*) desc
---14,642 patients with more than one colonoscpy (out of 138800)


select Patient_ID, MIN(admidate) admidate
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Patient_ID in (select Patient_ID
                                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                        where
droprecord is null
                                                                                        group by
patient_ID
                                                                                        having
COUNT(*) > 1)
        and dayspostopColonoscopyDiag between 180 and 365
group by Patient_ID
---1941


update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 0
where Patient_ID in (    select Patient_ID
                                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                        where
droprecord is null
                                                                                        group by
patient_ID
                                                                                        having
COUNT(*) > 1)
----(35084 row(s) affected)
```

```
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
                                                        (select Patient_ID, max(epistart) epistart
                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                        where Patient_ID in (    select Patient_ID
                                                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
        where droprecord is null
        group by patient_ID
        having COUNT(*) > 1 and MAX(dayspostopColonoscopyDiag) <= 180)
                                                        group by Patient_ID
                                                        ) t2
                                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                                        and
t1.epistart = t2.epistart

-----    (4352 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart
                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                        where Patient_ID in (select Patient_ID
                                                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
        where droprecord is null
        group by patient_ID
        having COUNT(*) > 1)
                                                        group by Patient_ID
                                                        ) t2
                                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                                        and
t1.epistart = t2.epistart
where t1.dayspostopColonoscopyDiag > 1826
-----(575 row(s) affected)
select  PAtient_ID, COUNT(*)
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Lastcolonoscopybeforediag = 1
group by Patient_ID
order by  COUNT(*) desc
---199 multiple rows
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Lastcolonoscopybeforediag = 1 and Patient_ID in (select  PAtient_ID
```

```
                                                    from
        Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where droprecord is null and
LastcolonoscopybeforeDiag = 1

                                                    group by Patient_ID

                                                    having  COUNT(*) > 1)
order by Patient_ID, admidate
---these all seem to be multiple episodes
select  PAtient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having  COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0
----9917
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0)

                                                                    and
dayspostopColonoscopyDiag between 181 and 365

                                                    group by Patient_ID

                                                    ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.epistart = t2.epistart
----(1944 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0)
```

```
                                                                    and
dayspostopColonoscopyDiag between 366 and 1096

                                                        group by Patient_ID

                                                        ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart
----(4447 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybefore Diag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart

                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                        where Patient_ID in (select Patient_ID

                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0)

                                                                    and
dayspostopColonoscopyDiag between 1097 and 1826

                                                        group by Patient_ID


                                                        ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart
---2355
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(Lastcolonoscopybeforediag)  desc
---there are 236 patients with multiple records for Lastcolonoscopybeforediag
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and patient_ID in (select Patient_ID
                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                    where
droprecord is null
                                                                    group by
Patient_ID
                                                                    having
sum(Lastcolonoscopybeforediag) > 1)
order by Patient_ID, admidate
----a lot of these do seem to be due to multiple episodes
---Lets first check that everyone has a last record
select Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having max(Lastcolonoscopybeforediag) = 0
---there are 1224 patietns here without one so lets look at these records first
```

```
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
Lastcolonoscopybeforediag, dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where PAtient_ID in (select Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where droprecord is null
                                                group by Patient_ID
                                                having max(Lastcolonoscopybeforediag) = 0        )

order by dayspostopColonoscopyDiag
--this is due to the patients having a record that more thn five years ago and also in the last 6
months
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart

                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                        where Patient_ID in (select Patient_ID


        from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        ---where droprecord is null


        group by Patient_ID


        having max(Lastcolonoscopybeforediag) = 0        )

                                                        group by Patient_ID


                                                        ) t2
                                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                                and
t1.epistart = t2.epistart
----    (62 row(s) affected)
---and back to the multiple records
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(Lastcolonoscopybeforediag)  desc
---there are 121 patients with multiple records for Lastcolonoscopybeforediag
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
Lastcolonoscopybeforediag, dayspostopColonoscopyDiag
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and patient_ID in (select Patient_ID
                                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                                where
droprecord is null
                                                                                                group by
Patient_ID
                                                                                                having
sum(Lastcolonoscopybeforediag) > 1)
order by Patient_ID, admidate
----a lot of these do seem to be due to multiple episodes
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 0
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  (select Patient_Id,
spell_Id, MIN(epiorder) minepi

                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                where Patient_ID in (select Patient_ID
```

```
                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                                    where
droprecord is null

                                                                    group by
Patient_ID

                                                                    having
sum(Lastcolonoscopybeforediag) >1)

                                                        and Lastcolonoscopybeforediag = 1

                                                group by  Patient_Id, spell_Id) t2
                                                on t1.Patient_ID = t2.Patient_ID
                                                and t1.SPELL_ID = t2.SPELL_ID
                                                and t1.EPIORDER = t2.minepi
-----(.190 row(s) affected)

select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag ,
colonoscopy_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where PAtient_ID in (select Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where droprecord is null
                                                group by Patient_ID
                                                having max(Lastcolonoscopybeforediag) = 0        )

order by Patient_ID      , ADMIDATE

---lets try updating these now

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join

                                                        (select Patient_ID, max(epistart) epistart,
max(SPELL_ID) spell

                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                        where droprecord is null and Patient_ID in (
        select Patient_ID

                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having max(Lastcolonoscopybeforediag) = 0 and MAX(dayspostopColonoscopyDiag) <= 180
        )

                                                        group by Patient_ID

                                                        ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart
                                                                    and
t1.SPELL_ID = t2.spell

------62+0
```

```
select  PAtient_ID, COUNT(*)
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Lastcolonoscopybeforediag = 1
group by Patient_ID
order by  COUNT(*) desc
---no  multiple rows
---LEts see how many rows there are left to update
select  PAtient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having  COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0
----13
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart, MIN(spell_ID) spell

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0)

                                                                        and
dayspostopColonoscopyDiag between 181 and 365

                                                    group by Patient_ID

                                                    ) t2
                                                                            on
t1.Patient_ID = t2.Patient_ID
                                                                            and
t1.epistart = t2.epistart
                                                                            and
t1.SPELL_ID = t2.spell
----(1 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart, MIN(spell_ID) spell_ID

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0)

                                                                        and
dayspostopColonoscopyDiag between 366 and 1096

                                                    group by Patient_ID
```

```
                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.epistart = t2.epistart
                                                                        and
t1.SPELL_ID = t2.spell_ID
----(6 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart, MIN(spell_ID)spell

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0)

                                                                    and
dayspostopColonoscopyDiag between 1097 and 1826

                                                    group by Patient_ID


                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.epistart = t2.epistart
                                                                        and
t1.SPELL_ID = t2.spell
---4
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(Lastcolonoscopybeforediag)  desc
---there are no  patients with multiple records for Lastcolonoscopybeforediag

---Lets now check that everyone has a last record
select Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having max(Lastcolonoscopybeforediag) = 0
---there are 7 patietns here without one so lets look at these records
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where PAtient_ID in (select Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where droprecord is null
                                                group by Patient_ID
                                                having max(Lastcolonoscopybeforediag) = 0        )
                and droprecord is null
order by Patient_ID, dayspostopColonoscopyDiag
---
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
```

```
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
min(dayspostopColonoscopyDiag) epistart, MIN(spell_ID)spell

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having  MAX(LastcolonoscopybeforeDiag) = 0)



                                                    group by Patient_ID


                                                    ) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    ---and
t1.dayspostopColonoscopyDiag = t2.epistart
                                                                                    and
t1.SPELL_ID = t2.spell
---(7 row(s) affected)
---Lets check these out again
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(Lastcolonoscopybeforediag)  desc
---again no multiples
select Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having max(Lastcolonoscopybeforediag) = 0


update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = 'Ascending colon'
where type_cancer = 'Ascending Colon'

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = 'Descending colon'
where type_cancer = 'Descending Colon'

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = 'Sigmoid colon'
where type_cancer = 'Sigmoid Colon'


update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set type_cancer = 'Transverse colon'
where type_cancer = 'Transverse Colon'
        /*
(8179 row(s) affected)

(2511 row(s) affected)

(15405 row(s) affected)

(5012 row(s) affected)

        */
```

```
---the following query looks at the length of time from colonoscopy to diagnosis

select Patient_ID, min(dayspostopColonoscopyDiag)
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and  IBD = 0
group by PAtient_ID
order by min(dayspostopColonoscopyDiag)

---and lets look at these by the type of diagnoses
select Patient_ID, min(dayspostopColonoscopyDiag)
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and  IBD = 0
group by PAtient_ID
having max( case  when ([dayspostop_ColonoscpyDiag_altered]<=180) and ([LastColonoscopybeforeDiag]=1)
and ([IBD]=0) then 'Diagnosed'
        when ([dayspostop_ColonoscpyDiag_altered]>180) and ([dayspostop_ColonoscpyDiag_altered]<=365)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then 'Definite_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>365) and ([dayspostop_ColonoscpyDiag_altered]<=1096)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then'Probable_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>1096) and
([dayspostop_ColonoscpyDiag_altered]<=1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'Latent'
when ([dayspostop_ColonoscpyDiag_altered]>1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'More5years' end) in ('Definite_Missed')
order by min(dayspostopColonoscopyDiag)
---probably missed
select Patient_ID, min(dayspostopColonoscopyDiag)
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and  IBD = 0
group by PAtient_ID
having max( case  when ([dayspostop_ColonoscpyDiag_altered]<=180) and ([LastColonoscopybeforeDiag]=1)
and ([IBD]=0) then 'Diagnosed'
        when ([dayspostop_ColonoscpyDiag_altered]>180) and ([dayspostop_ColonoscpyDiag_altered]<=365)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then 'Definite_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>365) and ([dayspostop_ColonoscpyDiag_altered]<=1096)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then'Probable_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>1096) and
([dayspostop_ColonoscpyDiag_altered]<=1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'Latent'
when ([dayspostop_ColonoscpyDiag_altered]>1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'More5years' end) in ('Probable_Missed')
order by min(dayspostopColonoscopyDiag)
---latent
select Patient_ID, min(dayspostopColonoscopyDiag)
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and  IBD = 0
group by PAtient_ID
having max( case  when ([dayspostop_ColonoscpyDiag_altered]<=180) and ([LastColonoscopybeforeDiag]=1)
and ([IBD]=0) then 'Diagnosed'
        when ([dayspostop_ColonoscpyDiag_altered]>180) and ([dayspostop_ColonoscpyDiag_altered]<=365)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then 'Definite_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>365) and ([dayspostop_ColonoscpyDiag_altered]<=1096)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then'Probable_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>1096) and
([dayspostop_ColonoscpyDiag_altered]<=1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'Latent'
when ([dayspostop_ColonoscpyDiag_altered]>1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'More5years' end) in ('Latent')
order by min(dayspostopColonoscopyDiag)

select Patient_ID, min(dayspostopColonoscopyDiag)
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and  IBD = 0
group by PAtient_ID
having max( case  when ([dayspostop_ColonoscpyDiag_altered]<=180) and ([LastColonoscopybeforeDiag]=1)
and ([IBD]=0) then 'Diagnosed'
        when ([dayspostop_ColonoscpyDiag_altered]>180) and ([dayspostop_ColonoscpyDiag_altered]<=365)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then 'Definite_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>365) and ([dayspostop_ColonoscpyDiag_altered]<=1096)
and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then'Probable_Missed'
        when ([dayspostop_ColonoscpyDiag_altered]>1096) and
([dayspostop_ColonoscpyDiag_altered]<=1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'Latent'
```

```
when ([dayspostop_ColonoscpyDiag_altered]>1826) and ([LastColonoscopybeforeDiag]=1) and ([IBD]=0) then
'More5years' end) in ('diagnosed', 'More5years')
order by min(dayspostopColonoscopyDiag)

----I'm slightly worried abour how I have defined the IBD status of the patient
set concat_null_yields_null off
select distinct t1.Patient_ID
from HED.dbo.HES_SPElls t1 inner join  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2
                                on t1.Patient_ID = t2.Patient_ID collate database_Default
                                and t1.admidate <= t2.firstadmissiondiagnosis
where
t1.DIAG_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12+t1.diag_13+t1.DIAG_14 like '%K5[01]%'
                and t2.IBD = 0 and t2.droprecord is null ;
---(1482 row(s) affected)

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set IBD = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2 inner join (select distinct
t1.Patient_ID

                                                                from HED.dbo.HES_SPElls t1 inner join
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2

                                                                                                on
t1.Patient_ID = t2.Patient_ID collate database_Default

                                                                                                and
t1.admidate <= t2.firstadmissiondiagnosis

                                                                where
t1.DIAG_01+t1.diag_02+t1.diag_03+t1.diag_04+t1.diag_05+t1.diag_06+t1.diag_07+t1.diag_08+t1.diag_09+t1.
diag_10+t1.diag_11+t1.diag_12+t1.diag_13+t1.DIAG_14 like '%K5[01]%'

                                                                                        and t2.IBD = 0 and
t2.droprecord is null) t1

        on t1.Patient_ID = t2.Patient_ID collate database_default ;
-----(1761 row(s) affected)



alter table Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
add lastScope datetime

update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set lastScope = t2.scopedate
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(colonoscopy_date) scopedate

                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                                where colonoscopy_date<=
Diag_date_altered

                                                                group by Patient_ID) t2

                on t1.patient_ID = t2.Patient_ID collate database_default
---(84030 row(s) affected)


----This script identifies the last colonoscopy
--first identify all the records with only one colonoscopy - Easy to define last
select Patient_ID, COUNT(*)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by patient_ID
Order by COUNT(*) desc
---14,429 patients with one colonoscpy (out of 74429)

select Patient_ID, MIN(admidate) admidate
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Patient_ID in (select Patient_ID
```

```
                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                    where
droprecord is null
                                                                    group by
patient_ID
                                                                    having
COUNT(*) > 1)
        and dayspostopColonoscopyDiag between 180 and 365
group by Patient_ID
---860


update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag =0
where Patient_ID in (    select Patient_ID
                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                    where
droprecord is null
                                                                    group by
patient_ID
                                                                    having
COUNT(*) > 1)
----(15811 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join
                                                    (select Patient_ID, max(epistart) epistart
                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                    where Patient_ID in (    select Patient_ID
                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

        where droprecord is null

        group by patient_ID

        having COUNT(*) > 1 and MAX(dayspostopColonoscopyDiag) <= 180)
                                                    group by Patient_ID
                                                    ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart

-----    (2334 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart
                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                    where Patient_ID in (select Patient_ID
                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

        where droprecord is null
```

```
        group by patient_ID

        having COUNT(*) > 1)
                                                    group by Patient_ID
                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.epistart = t2.epistart
where t1.dayspostopColonoscopyDiag > 1826
-----(748 row(s) affected)
select  PAtient_ID, COUNT(*)
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and LastcolonoscopybeforeDiag = 1
group by Patient_ID
order by  COUNT(*) desc
---104 multiple rows
select *
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and LastcolonoscopybeforeDiag = 1 and Patient_ID in (select  PAtient_ID
                                                    from
        Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                    where droprecord is null and
LastcolonoscopybeforeDiag = 1
                                                    group by Patient_ID
                                                    having  COUNT(*) > 1)
order by Patient_ID, admidate
---these all seem to be multiple episodes
select  PAtient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having  COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0
----3944
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart
                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                    where Patient_ID in (select Patient_ID
                                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

        where droprecord is null

        group by patient_ID

        having COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0)
                                                                        and
dayspostopColonoscopyDiag between 181 and 365
                                                    group by Patient_ID
                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
```

```
                                                                    and
t1.epistart = t2.epistart
----(857 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0)

                                                                    and
dayspostopColonoscopyDiag between 366 and 1096

                                                    group by Patient_ID

                                                    ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart
----(1887 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart

                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                    where Patient_ID in (select Patient_ID

                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        where droprecord is null


        group by patient_ID


        having COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0)

                                                                    and
dayspostopColonoscopyDiag between 1097 and 1826

                                                    group by Patient_ID


                                                    ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart
---1161
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
```

```
Order by sum(Lastcolonoscopybeforediag)  desc
---there are 121 patients with multiple records for Lastcolonoscopybeforediag
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and patient_ID in (select Patient_ID
                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                where
droprecord is null
                                                                                group by
Patient_ID
                                                                                having
sum(Lastcolonoscopybeforediag) > 1)
order by Patient_ID, admidate
----a lot of these do seem to be due to multiple episodes
---Lets first check that everyone has a last record
select Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having max(Lastcolonoscopybeforediag) = 0
---there are 57 patietns here without one so lets look at these records first
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where PAtient_ID in (select Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where droprecord is null
                                                group by Patient_ID
                                                having max(Lastcolonoscopybeforediag) = 0        )

order by dayspostopColonoscopyDiag
--this is due to the patients having a record that more thn five years ago and also in the last 6
months
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart

                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                        where Patient_ID in (select Patient_ID


        from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses


        ---where droprecord is null


        group by Patient_ID


        having max(Lastcolonoscopybeforediag) = 0        )

                                                        group by Patient_ID


                                                        ) t2
                                                                                on
t1.Patient_ID = t2.Patient_ID
                                                                                and
t1.epistart = t2.epistart
----    (57 row(s) affected)
---and back to the multiple records
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(Lastcolonoscopybeforediag)  desc
---there are 121 patients with multiple records for Lastcolonoscopybeforediag
```

```
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag
from  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and patient_ID in (select Patient_ID
                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                                                where
droprecord is null
                                                                                group by
Patient_ID
                                                                                having
sum(Lastcolonoscopybeforediag) > 1)
order by Patient_ID, admidate
----a lot of these do seem to be due to multiple episodes
update Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 0
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join  (select Patient_Id,
spell_Id, MIN(epiorder) minepi

                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                where Patient_ID in (select Patient_ID

                                                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                                                where
droprecord is null

                                                                                group by
Patient_ID

                                                                                having
sum(Lastcolonoscopybeforediag) >1)

                                                        and Lastcolonoscopybeforediag = 1

                                                group by  Patient_Id, spell_Id) t2
                                                on t1.Patient_ID = t2.Patient_ID
                                                and t1.SPELL_ID = t2.SPELL_ID
                                                and t1.EPIORDER = t2.minepi
-----(190 row(s) affected)

select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag ,
colonoscopy_date
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where PAtient_ID in (select Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where droprecord is null
                                                group by Patient_ID
                                                having max(Lastcolonoscopybeforediag) = 0        )

order by Patient_ID      , ADMIDATE

---lets try updating these now

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set LastcolonoscopybeforeDiag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join

                                                        (select Patient_ID, max(epistart) epistart,
max(SPELL_ID) spell

                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

                                                        where Patient_ID in (    select Patient_ID

                                                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
```

```
        where droprecord is null

        group by patient_ID

        having max(Lastcolonoscopybeforediag) = 0 and MAX(dayspostopColonoscopyDiag) <= 180
        )
                                                    group by Patient_ID
                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.epistart = t2.epistart
                                                                        and
t1.SPELL_ID = t2.spell

------55

update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart, MIN(spell_ID) spell
                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                    where Patient_ID in (select Patient_ID
                                                                                        from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

        where droprecord is null

        group by patient_ID

        having max(Lastcolonoscopybeforediag) = 0)
                                                    group by Patient_ID
                                                    ) t2
                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                        and
t1.epistart = t2.epistart
                                                                        and
t1.SPELL_ID = t2.spell
where t1.dayspostopColonoscopyDiag > 1826
-----(0 row(s) affected)
select  PAtient_ID, COUNT(*)
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null and Lastcolonoscopybeforediag = 1
group by Patient_ID
order by  COUNT(*) desc
---no  multiple rows
---LEts see how many rows there are left to update
select  PAtient_ID
from    Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having  COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0
----13
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart, MIN(spell_ID) spell
```

```
                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                    where Patient_ID in (select Patient_ID
                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
        where droprecord is null
        group by patient_ID
        having COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0)
                                                                    and
dayspostopColonoscopyDiag between 181 and 365
                                                    group by Patient_ID
                                                    ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart
                                                                    and
t1.SPELL_ID = t2.spell
----(1 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart, MIN(spell_ID) spell_ID
                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                    where Patient_ID in (select Patient_ID
                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
        where droprecord is null
        group by patient_ID
        having COUNT(*) > 1 and MAX(Lastcolonoscopybeforediag) = 0)
                                                                    and
dayspostopColonoscopyDiag between 366 and 1096
                                                    group by Patient_ID
                                                    ) t2
                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                    and
t1.epistart = t2.epistart
                                                                    and
t1.SPELL_ID = t2.spell_ID
----(6 row(s) affected)
update  Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
set Lastcolonoscopybeforediag = 1
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t1 inner join (select Patient_ID,
max(epistart) epistart, MIN(spell_ID)spell
                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
```

```
                                                    where Patient_ID in (select Patient_ID
                                                                                                                    from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses

        where droprecord is null

        group by patient_ID

        having COUNT(*) > 1 and MAX(LastcolonoscopybeforeDiag) = 0)
                                                                                    and
dayspostopColonoscopyDiag between 1097 and 1826
                                                    group by Patient_ID

                                                    ) t2
                                                                                        on
t1.Patient_ID = t2.Patient_ID
                                                                                        and
t1.epistart = t2.epistart
                                                                                        and
t1.SPELL_ID = t2.spell
---4
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(Lastcolonoscopybeforediag)  desc
---there are no  patients with multiple records for Lastcolonoscopybeforediag

---Lets now check that everyone has a last record
select Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having max(Lastcolonoscopybeforediag) = 0
---there are 2 patietns here without one so lets look at these records
select Patient_ID, SPELL_ID, SPELL_FINYR, EPISTART, EPIORDER, ADMIDATE, DISDATE,
LastcolonoscopybeforeDiag, dayspostopColonoscopyDiag
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where PAtient_ID in (select Patient_ID
                                                from
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
                                                where droprecord is null
                                                group by Patient_ID
                                                having max(Lastcolonoscopybeforediag) = 0        )

order by dayspostopColonoscopyDiag
---LEts update these manually

---Lets check these out again
select Patient_ID, sum(Lastcolonoscopybeforediag)
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
Order by sum(Lastcolonoscopybeforediag)  desc
---again no multiples
select Patient_ID
from Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses
where droprecord is null
group by Patient_ID
having max(Lastcolonoscopybeforediag) = 0
---and everyone has a last record


SET CONCAT_NULL_YIELDS_NULL OFF
select *
into Cancer_research.dbo.FE_20131018_AllColonoscopies
```

```
from (select Patient_ID, SPELL_ID, SPELL_FINYR, EPIORDER,EPISTART, ADMIDATE, DISDATE, ADMIMETH,
CLASSPAT, DIAG_01, DIAG_02, DIAG_03, DIAG_04,DIAG_05, DIAG_06,
                DIAG_07, DIAG_08,DIAG_09, DIAG_10,DIAG_11, DIAG_12, DIAG_13, DIAG_14, DISMETH, GPPRAC,
MAINSPEF, OPERTN_01, OPERTN_02, OPERTN_03, OPERTN_04, OPERTN_05,
                OPERTN_06, OPERTN_07, opertn_08, OPERTN_09, OPERTN_10, OPERTN_11, OPERTN_12,
OPERTN_13, OPERTN_14, LEFT(procode,3) procode, SEX, STARTAGE, quintiles , resgor,
                case when opertn_01 like '%H20[16]%' OR OPERTN_01 like '%H22%' then OPDATE_01
                        when opertn_02 like '%H20[16]%' OR OPERTN_02 like '%H22%' then OPDATE_02
                        when opertn_03 like '%H20[16]%' OR OPERTN_03 like '%H22%' then OPDATE_03
                        when opertn_04 like '%H20[16]%' OR OPERTN_04 like '%H22%' then OPDATE_04
                        when opertn_05 like '%H20[16]%' OR OPERTN_05 like '%H22%' then OPDATE_05
                        when opertn_06 like '%H20[16]%' OR OPERTN_06 like '%H22%' then OPDATE_06
                        when opertn_07 like '%H20[16]%' OR OPERTN_07 like '%H22%' then OPDATE_07
                        when opertn_08 like '%H20[16]%' OR OPERTN_08 like '%H22%' then OPDATE_08
                        when opertn_09 like '%H20[16]%' OR OPERTN_09 like '%H22%' then OPDATE_09
                        when opertn_10 like '%H20[16]%' OR OPERTN_10 like '%H22%' then OPDATE_10
                        when opertn_11 like '%H20[16]%' OR OPERTN_11 like '%H22%' then OPDATE_11
                        when opertn_12 like '%H20[16]%' OR OPERTN_12 like '%H22%' then OPDATE_12
                        when opertn_13 like '%H20[16]%' OR OPERTN_13 like '%H22%' then OPDATE_13
                        when opertn_14 like '%H20[16]%' OR OPERTN_14 like '%H22%' then OPDATE_14
                        when opertn_15 like '%H20[16]%' OR OPERTN_15 like '%H22%' then OPDATE_15
                        when opertn_16 like '%H20[16]%' OR OPERTN_16 like '%H22%' then OPDATE_16
                        when opertn_17 like '%H20[16]%' OR OPERTN_17 like '%H22%' then OPDATE_17
                        when opertn_18 like '%H20[16]%' OR OPERTN_18 like '%H22%' then OPDATE_18
                        when opertn_19 like '%H20[16]%' OR OPERTN_19 like '%H22%' then OPDATE_19
                        when opertn_20 like '%H20[16]%' OR OPERTN_20 like '%H22%' then OPDATE_20 END
as colonoscopy_date

                from HED.dbo.HES_EPISODES_part1
                where (opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+ OPERTN_06+opertn_07+
OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+ OPERTN_16+opertn_17+
OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24 like '%H20[16]%' or
                                opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+
OPERTN_06+opertn_07+ OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+
OPERTN_16+opertn_17+ OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24   like
'%H22%')
                                                and ADMIDATE > '20030331'
                union ALL
                select Patient_ID, SPELL_ID, SPELL_FINYR, EPIORDER,EPISTART,  ADMIDATE, DISDATE,
ADMIMETH, CLASSPAT, DIAG_01, DIAG_02, DIAG_03, DIAG_04,DIAG_05, DIAG_06,
                DIAG_07, DIAG_08,DIAG_09, DIAG_10,DIAG_11, DIAG_12, DIAG_13, DIAG_14, DISMETH, GPPRAC,
MAINSPEF, OPERTN_01, OPERTN_02, OPERTN_03, OPERTN_04, OPERTN_05,
                OPERTN_06, OPERTN_07, opertn_08, OPERTN_09, OPERTN_10, OPERTN_11, OPERTN_12,
OPERTN_13, OPERTN_14, LEFT(procode,3) procode, SEX, STARTAGE, quintiles , resgor
                , case when opertn_01 like '%H20[16]%' OR OPERTN_01 like '%H22%' then OPDATE_01
                        when opertn_02 like '%H20[16]%' OR OPERTN_02 like '%H22%' then OPDATE_02
                        when opertn_03 like '%H20[16]%' OR OPERTN_03 like '%H22%' then OPDATE_03
                        when opertn_04 like '%H20[16]%' OR OPERTN_04 like '%H22%' then OPDATE_04
                        when opertn_05 like '%H20[16]%' OR OPERTN_05 like '%H22%' then OPDATE_05
                        when opertn_06 like '%H20[16]%' OR OPERTN_06 like '%H22%' then OPDATE_06
                        when opertn_07 like '%H20[16]%' OR OPERTN_07 like '%H22%' then OPDATE_07
                        when opertn_08 like '%H20[16]%' OR OPERTN_08 like '%H22%' then OPDATE_08
                        when opertn_09 like '%H20[16]%' OR OPERTN_09 like '%H22%' then OPDATE_09
                        when opertn_10 like '%H20[16]%' OR OPERTN_10 like '%H22%' then OPDATE_10
                        when opertn_11 like '%H20[16]%' OR OPERTN_11 like '%H22%' then OPDATE_11
                        when opertn_12 like '%H20[16]%' OR OPERTN_12 like '%H22%' then OPDATE_12
                        when opertn_13 like '%H20[16]%' OR OPERTN_13 like '%H22%' then OPDATE_13
                        when opertn_14 like '%H20[16]%' OR OPERTN_14 like '%H22%' then OPDATE_14
                        when opertn_15 like '%H20[16]%' OR OPERTN_15 like '%H22%' then OPDATE_15
                        when opertn_16 like '%H20[16]%' OR OPERTN_16 like '%H22%' then OPDATE_16
                        when opertn_17 like '%H20[16]%' OR OPERTN_17 like '%H22%' then OPDATE_17
                        when opertn_18 like '%H20[16]%' OR OPERTN_18 like '%H22%' then OPDATE_18
                        when opertn_19 like '%H20[16]%' OR OPERTN_19 like '%H22%' then OPDATE_19
                        when opertn_20 like '%H20[16]%' OR OPERTN_20 like '%H22%' then OPDATE_20 END
as colonoscopy_date
                from HED.dbo.HES_EPISODES
                where (opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+ OPERTN_06+opertn_07+
OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+ OPERTN_16+opertn_17+
OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24 like '%H20[16]%' or
                                                opertn_01+OPERTN_02+opertn_03+OPERTN_04+opertn_05+
OPERTN_06+opertn_07+ OPERTN_08+ opertn_09+OPERTN_10+opertn_11+OPERTN_12+opertn_13+OPERTN_14+opertn_15+
OPERTN_16+opertn_17+ OPERTN_18+ opertn_19+OPERTN_20+opertn_21+OPERTN_22 +opertn_23+OPERTN_24   like
'%H22%')
                                                and ADMIDATE < '20090401') t1
```

```
-----(1468549 row(s) affected)

select *
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2
                                                        on t1.Patient_ID = t2.Patient_ID
where t1.admidate > t2.firstAdmissionDiagnosis
---22641 rows
---and now deleting them
delete t1
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2
                                                        on t1.Patient_ID = t2.Patient_ID
where t1.ADMIDATE > t2.firstAdmissionDiagnosis
----(20854 row(s) affected)

-------------------------------------------------cleaning!-------------------------------------
---------------------------

-----Apply the same cleaning rules as those in the full data tavle
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set DISDATE = t2.disdate
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join HED.dbo.HES_SPells t2
                                                    on t1.Patient_ID = t2.PATIENT_ID
                                                    and t1.SPELL_ID = t2.SPELL_ID
where t1.DISDATE is null
------(47426 row(s) affected)

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set colonoscopy_date  = disdate
where colonoscopy_date not between ADMIDATE and  disdate
----(3548 row(s) affected)

alter table Cancer_research.dbo.FE_20131018_AllColonoscopies
add droprecord int;

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord = 1
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2
                                                        on t1.Patient_ID = t2.Patient_ID
where datediff(dd, t1.colonoscopy_date,  t2.firstAdmissionDiagnosis)  < -90
-----
---check that this has worked...
select t1.Patient_ID, t1.colonoscopy_date, t2.firstAdmissionDiagnosis, t1.droprecord
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join
Cancer_research.dbo.FE_20131011_ColonoscopyMissedDiagnoses t2
                                                        on t1.Patient_ID = t2.Patient_ID
where t1.droprecord = 1
---It apopears to have done!
---updating the discharge method to that of the spell
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set DISMETH = t2.dismeth
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join HED.dbo.HES_SPElls t2
                                                                on t1.Patient_ID =
t2.PATIENT_ID
                                                                and t1.SPELL_ID =
t2.SPELL_ID

----(1446906 row(s) affected)

select sex, COUNT(patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
group by sex

---some 0's and 9's
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord =1
where SEX in (0,9)
---     (151 row(s) affected)
----lets look at the region of residence
select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
```

```
group by resgor
order by resgor

---- no resgor, 6k with an invlaid resgor, 6k with a welsh/scottish
select resgor, COUNT(distinct Patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where Patient_ID in (select Patient_ID
                                        from Cancer_research.dbo.FE_20131018_AllColonoscopies
                                        where resgor in ('u','x','y','z')
                                        )
         and resgor in ('a','b','c','d','e','f','g','h','i','k','j')
group by resgor
order by resgor
-----theres quite a few upodates here....
select distinct Patient_ID
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where Patient_ID in (select Patient_ID
                                        from Cancer_research.dbo.FE_20131018_AllColonoscopies
                                        where resgor in ('u','x','y','z')
                                        )
         and resgor in ('a','b','c','d','e','f','g','h','i','k','j')
 ----1062

 select *
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where Patient_ID in (select distinct Patient_ID
                                                from Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                where Patient_ID in (select Patient_ID
                                                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                                                        where resgor
in ('u','y','z')
                                                                                                        )
                                                                                and resgor in
('a','b','c','d','e','f','g','h','i','k','j'))
order by Patient_ID, admidate
---lets leave the 'u's as they are and update the y's and z's
select distinct t1.PAtient_ID, t1.resgor, t2.resgor
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct Patient_ID,
resgor

                                        from  Cancer_research.dbo.FE_20131018_AllColonoscopies

                                        where resgor in
('a','b','c','d','e','f','g','h','i','k','j')) t2
                                                                                                        on
t1.patient_ID = t2.patient_ID
where t1.resgor in ('y','z')
order by PAtient_ID
--979
select distinct patient_ID
from (select distinct t1.PAtient_ID, t1.resgor, t2.resgor resgornew
                        from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select
distinct Patient_ID, resgor

                                                                                from
Cancer_research.dbo.FE_20131018_AllColonoscopies

                                                                                where resgor in
('a','b','c','d','e','f','g','h','i','k','j')) t2

                on t1.patient_ID = t2.patient_ID
                        where t1.resgor in ('y','z'))  a
---956
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor  = t2.resgornew
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.PAtient_ID,
t1.resgor, t2.resgor resgornew

                                                from Cancer_research.dbo.FE_20131018_AllColonoscopies
t1 inner join (select distinct Patient_ID, resgor
```

```
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies

                                                                    where resgor
in ('a','b','c','d','e','f','g','h','i','k','j')) t2

                                        on t1.patient_ID = t2.patient_ID

                                                where t1.resgor in ('y','z')) t2
                                                                    on t1.Patient_ID=
t2.PAtient_ID
                                                                    and t1.resgor =
t2.resgor
----(388 row(s) affected)
select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord  is null
group by resgor
order by resgor
---there are still about 5k patients with an invalid GOR code
---lets check these patients back to the HES tables
select t1.Patient_ID, t1.spell_ID, t1.resgor, t2.resgor
from HED.dbo.HES_Episodes t1 inner join (select resgor, Patient_ID, spell_ID
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                    where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
                                                    and t1.SPELL_ID = t2.SPELL_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---11 rows - lets update
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor  = t2.resgornew
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join      (select t1.Patient_ID,
t1.spell_ID, t1.resgor resgornew, t2.resgor

                        from HED.dbo.HES_Episodes t1 inner join (select resgor, Patient_ID, spell_ID

        from Cancer_research.dbo.FE_20131018_AllColonoscopies

        where droprecord  is null and resgor in ('y','z')) t2

                                                                    on t1.PATIENT_ID =
t2.Patient_ID

                                                                    and t1.SPELL_ID =
t2.SPELL_ID

                        where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                    on  t1.Patient_ID=
t2.PAtient_ID
                                                                    and t1.resgor =
t2.resgor
----(8 row(s) affected)
select t1.Patient_ID, t1.spell_ID, t1.resgor, t2.resgor
from HED.dbo.HES_Episodes_part1 t1 inner join (select resgor, Patient_ID, spell_ID
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                    where droprecord  is
null and resgor in ('y','z')) t2
                                                on t1.PATIENT_ID = t2.Patient_ID
                                                    and t1.SPELL_ID = t2.SPELL_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
----13 rows
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor  = t2.resgornew
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join      (select t1.Patient_ID,
t1.spell_ID, t1.resgor resgornew, t2.resgor
```

```
                        from HED.dbo.HES_Episodes_part1 t1 inner join (select resgor, Patient_ID,
spell_ID

        from Cancer_research.dbo.FE_20131018_AllColonoscopies

        where droprecord  is null and resgor in ('y','z')) t2

                                                                    on t1.PATIENT_ID =
t2.Patient_ID

                                                                        and t1.SPELL_ID =
t2.SPELL_ID

                        where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                    on  t1.Patient_ID=
t2.PAtient_ID

                                                                        and t1.resgor =
t2.resgor
---(5 row(s) affected
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes_part1 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                        where droprecord  is
null and resgor in ('y','z')) t2
                                            on t1.PATIENT_ID = t2.Patient_ID
                                                    and abs(DATEDIFF(dd, t1.Admidate,
t2.Admidate)) between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
----97 rows

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.Patient_ID,
t1.resgor resgornew, t2.resgor

                                            from HED.dbo.HES_Episodes_part1 t1 inner join (select
resgor, Patient_ID, Admidate

                                from Cancer_research.dbo.FE_20131018_AllColonoscopies

                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                    on
t1.PATIENT_ID = t2.Patient_ID

        and abs(DATEDIFF(dd, t1.Admidate, t2.Admidate)) between 0 and 90

                                            where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                        and t1.resgor =
t2.resgor
----(103 row(s) affected)

select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                        where droprecord  is
null and resgor in ('y','z')) t2
                                            on t1.PATIENT_ID = t2.Patient_ID
                                                    and abs(DATEDIFF(dd, t1.Admidate,
t2.Admidate)) between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')

----286
```

```
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.Patient_ID,
t1.resgor resgornew, t2.resgor

                                                from HED.dbo.HES_Episodes t1 inner join (select
resgor, Patient_ID, Admidate

                                    from Cancer_research.dbo.FE_20131018_AllColonoscopies

                                    where droprecord  is null and resgor in ('y','z')) t2

                                                                                                    on
t1.PATIENT_ID = t2.Patient_ID

        and abs(DATEDIFF(dd, t1.Admidate, t2.Admidate)) between 0 and 90

                                                where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                on t1.Patient_ID =
t2.Patient_ID
                                                                                and t1.resgor =
t2.resgor
----(294 row(s) affected)

select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord  is null
group by resgor
order by resgor
----down to 4.5k
---lets make a start on OP
select distinct t1.PatientID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_200708 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENTID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
-----264
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.PatientID,
t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_200708 t1 inner join (select resgor,
Patient_ID, Admidate

                from Cancer_research.dbo.FE_20131018_AllColonoscopies

                where droprecord  is null and resgor in ('y','z')) t2

                                                                                on t1.PATIENTID =
t2.Patient_ID

                                                                                        and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                on t1.Patient_ID =
t2.PatientID
                                                                                and t1.resgor =
t2.resgor;
----(271 row(s) affected)
select distinct t1.PatientID, t1.resgor resgornew, t2.resgor
```

```
from HES_DW.dbo.HES_OP_200607 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENTID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
-----46 rows
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.PatientID,
t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_200607 t1 inner join (select resgor,
Patient_ID, Admidate


                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                where droprecord  is null and resgor in ('y','z')) t2

                                                                                on t1.PATIENTID =
t2.Patient_ID

                                                                                        and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                on t1.Patient_ID =
t2.PatientID
                                                                                and t1.resgor =
t2.resgor;
----(49 row(s) affected
select distinct t1.PatientID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_200809 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENTID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
-----70 rows
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.PatientID,
t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_200809 t1 inner join (select resgor,
Patient_ID, Admidate


                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                where droprecord  is null and resgor in ('y','z')) t2

                                                                                on t1.PATIENTID =
t2.Patient_ID

                                                                                        and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                on t1.Patient_ID =
t2.PatientID
                                                                                and t1.resgor =
t2.resgor;
```

```
----(71 row(s) affected)

select distinct t1.PatientID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_200910 t1 inner join (select resgor, Patient_ID, Admidate
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                    where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENTID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
-----69 rows
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.PatientID,
t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_200910 t1 inner join (select resgor,
Patient_ID, Admidate


                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                where droprecord  is null and resgor in ('y','z')) t2

                                                                    on t1.PATIENTID =
t2.Patient_ID

                                                                            and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                    on t1.Patient_ID =
t2.PatientID
                                                                    and t1.resgor =
t2.resgor;
-----(70 row(s) affected)

select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_201011 t1 inner join (select resgor, Patient_ID, Admidate
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                    where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENT_ID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
-----86 rows
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.Patient_ID,
t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_201011 t1 inner join (select resgor,
Patient_ID, Admidate


                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                where droprecord  is null and resgor in ('y','z')) t2

                                                                    on t1.PATIENT_ID =
t2.Patient_ID

                                                                            and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
```

```
                                                                    on t1.Patient_ID =
t2.Patient_ID
                                                                    and t1.resgor =
t2.resgor;

----(92 row(s) affected)
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_201112 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENT_ID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
-----79 rows
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.Patient_ID,
t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_201112 t1 inner join (select resgor,
Patient_ID, Admidate


                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                where droprecord  is null and resgor in ('y','z')) t2

                                                                    on t1.PATIENT_ID =
t2.Patient_ID

                                                                                and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90

                                        where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                    on t1.Patient_ID =
t2.Patient_ID
                                                                    and t1.resgor =
t2.resgor;
----(81 row(s) affected)
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HES_DW.dbo.HES_OP_201213 t1 inner join (select resgor, Patient_ID, Admidate
                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                        where droprecord  is null
and resgor in ('y','z')) t2
                                        on t1.PATIENT_ID = t2.Patient_ID
                                                and abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate))
between 0 and 90
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
-----5 rows
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.Patient_ID,
t1.resgor resgornew, t2.resgor

                                        from HES_DW.dbo.HES_OP_201213 t1 inner join (select resgor,
Patient_ID, Admidate


                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                where droprecord  is null and resgor in ('y','z')) t2

                                                                    on t1.PATIENT_ID =
t2.Patient_ID

                                                                                and
abs(DATEDIFF(dd, t1.Apptdate, t2.Admidate)) between 0 and 90
```

```
                                    where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                    on t1.Patient_ID =
t2.Patient_ID
                                                                    and t1.resgor =
t2.resgor;
-----(5 row(s) affected)
---check how we are doing
select resgor, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord  is null
group by resgor
order by resgor
---about 4k left
---Just check all episodes for an update
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes t1 inner join (select resgor, Patient_ID, Admidate
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                    where droprecord  is
null and resgor in ('y','z')) t2
                                            on t1.PATIENT_ID = t2.Patient_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---819 rows
select distinct t1.Patient_ID, t1.resgor resgornew, t2.resgor
from HED.dbo.HES_Episodes_Part1 t1 inner join (select resgor, Patient_ID, Admidate
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                    where droprecord  is
null and resgor in ('y','z')) t2
                                            on t1.PATIENT_ID = t2.Patient_ID
where t1.RESGOR in ('a','b','c','d','e','f','g','h','i','j','k')
---710

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.Patient_ID,
t1.resgor resgornew, t2.resgor

                                            from HED.dbo.HES_Episodes t1 inner join (select
resgor, Patient_ID, Admidate


                                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                    on
t1.PATIENT_ID = t2.Patient_ID



                                            where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                    on t1.Patient_ID =
t2.Patient_ID
                                                                    and t1.resgor =
t2.resgor

----(837 row(s) affected)

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set resgor = t2.resgornew
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select distinct t1.Patient_ID,
t1.resgor resgornew, t2.resgor

                                            from HED.dbo.HES_Episodes_part1 t1 inner join (select
resgor, Patient_ID, Admidate


                                from Cancer_research.dbo.FE_20131018_AllColonoscopies
```

```
                                where droprecord  is null and resgor in ('y','z')) t2

                                                                                                    on
t1.PATIENT_ID = t2.Patient_ID



                                                    where t1.RESGOR in
('a','b','c','d','e','f','g','h','i','j','k')) t2
                                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                                        and t1.resgor =
t2.resgor

---211



update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord = 1
where Patient_ID in (select distinct Patient_ID
                                        from  Cancer_research.dbo.FE_20131018_AllColonoscopies
                                where resgor in ('s','w','x'))

----(8999 row(s) affected)
----check the rest of the resgors
select resgor, COUNT(patient_ID)
from  Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord is null
group by resgor
---there are about 2k now still without a valid resgor - I shall drop theser
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord = 1
where resgor in ('y', 'z') and droprecord is null

----(1716 row(s) affected)



---check the spell year

select SPELL_FINYR, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord is null
group by SPELL_FINYR
order by SPELL_FINYR
---there are 1090 patients where the spell year is 9999 and they currently have not been dropped
select *
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where SPELL_FINYR   = '9999'
Order by Patient_ID, ADMIDATE
---1355 rows
---lets link these back to the hes tables

select t2.*
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select Patient_ID, spell_ID,
epiorder, admidate, disdate, epistart, epiend, epistat, SPELL_FINYR

                                        from HED.dbo.HES_EPISODES

                                union all

                                        select Patient_ID, spell_ID, epiorder, admidate, disdate,
epistart, epiend, epistat, SPELL_FINYR

                                        from HED.dbo.HES_EPISODES_Part1) t2
                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                    and
t1.SPELL_ID = t2.SPELL_ID
                                                                                    and
t1.EPIORDER = t2.Epiorder
```

```
where t1.SPELL_FINYR   = '9999'
Order by Patient_ID, ADMIDATE
---1355 rows
---all these episodes have finished so I'm not sure what has gone on - it is something to do with the
date of discharge, lets look to see whether there are any epsiodes in the same spell with a
---date of discharge
select t2.*
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select Patient_ID, spell_ID,
epi_last, epiorder, admidate, disdate, epistart, epiend, epistat, DISMETH, SPELL_FINYR

                                        from HED.dbo.HES_EPISODES

                                union all

                                        select Patient_ID, spell_ID, epiorder, epi_last,  admidate,
disdate, epistart, epiend, epistat, DISMETH,SPELL_FINYR

                                        from HED.dbo.HES_EPISODES_Part1) t2
                                                                                                    on
t1.Patient_ID = t2.Patient_ID
                                                                                                    and
t1.SPELL_ID = t2.SPELL_ID

where t1.SPELL_FINYR   = '9999'
Order by dismeth
---probably best to remove these records, for the most part it looks as if these spells have not
finished
--16348
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord = 1
where SPELL_FINYR = '9999'
---(1355 row(s) affected)

-----Look at admimeth
select admimeth, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord  is null
group by admimeth
order by admimeth

---look to see whether there is any difference between this and the spells table
select t1.Patient_ID, t2.spell_ID, t1.Admimeth, t2.Admimeth
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join HED.dbo.HES_SPELLS t2
                                                                                                    on
t1.Patient_ID = t2.PATIENT_ID
                                                                                                    and
t1.SPELL_ID = t2.Spell_ID
where t1.ADMIMETH > 90 and t2.ADMIMETH <99
---no...



---Now using the episodes table
select t1.Patient_ID, t1.Admimeth, t2.Admimeth
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join HED.dbo.HES_Episodes_part1 t2
                                                                                                    on
t1.Patient_ID = t2.PATIENT_ID
                                                                                                    and
t1.SPELL_ID = t2.Spell_ID
where t1.ADMIMETH > 90 and t2.ADMIMETH <99
----not sure what to do with these records -  I don't suppose it will matter too much
---check disdate
select disdate, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord  is null
group by disdate
order by disdate
--- there are three nulls here
select t1.Patient_ID, t1.SPELL_ID, t2.Disdate
from  HED.dbo.HES_SPELLS t2 inner join (select t1.Patient_ID, t1.SPELL_ID
                                                                                        from
Cancer_research.dbo.FE_20131018_AllColonoscopies t1
                                                                                        where droprecord  is
null and t1.DISDATE is null) t1
```

```
                                                    on t1.SPELL_ID = t2.SPELL_ID
                                                    and t1.Patient_ID = t2.PATIENT_ID

---these patients and spells are not in the spells table
select t1.Patient_ID, t1.SPELL_ID, t2.Disdate
from  HED.dbo.HES_EPISODES t2 inner join (select t1.Patient_ID, t1.SPELL_ID
                                                                    from
Cancer_research.dbo.FE_20131018_AllColonoscopies t1
                                                                    where droprecord  is
null and t1.DISDATE is null) t1
                                                    on t1.SPELL_ID = t2.SPELL_ID
                                                    and t1.Patient_ID = t2.PATIENT_ID

----The disdates are null so maybe the spell hasn't finished so lets drop these


---move onto dismeth
select dismeth, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord  is null
group by dismeth
order by dismeth
---there are 471 here with validation errors - I'm not sure what to do with these either!
---lets look at the class pats
select classpat, COUNT(distinct patient_ID)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord  is null
group by classpat
order by classpat
---we've got some regular day attenders here ....
---lets loook at these patients
select *
from  Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord is null and CLASSPAT =3
order by Patient_ID
---2636 rows



----Lets add dod and causes of death to this table
alter table Cancer_research.dbo.FE_20131018_AllColonoscopies
add dod datetime

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set dod = t2.DOD
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select Extract_HeSID, MAX(DOD)
DOD

                                        from HES_DW_ONS.dbo.HED_ONS_Mortality

                                        group by EXTRACT_HESID) t2

        on t1.Patient_ID = t2.EXTRACT_HESID collate DATABASE_default
----(435958 row(s) affected)


---check whether there are any deaths recorded as happening before the first record of GI cancer
select Patient_ID, SPELL_FINYR, ADMIDATE, DISDATE,dismeth,  DOD, DATEDIFF(dd, disdate, dod),
DATEDIFF(dd, admidate, dod)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where DOD is not null    and DOD < DISDATE
order by  DATEDIFF(dd, disdate, dod)
--- 729 rows, and most of these the DOD is a  long time before the cancer aignosis
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord = 1
where  DOD is not null   and DOD < DISDATE and DOD < admidate
-----(392 row(s) affected)

--lets just check the ages of these paitnets
select startAge, COUNT(distinct Patient_ID)
from (select Patient_ID, MIN(startAge) StartAgE
                from Cancer_research.dbo.FE_20131018_AllColonoscopies
                group by Patient_ID) t1
group by StartAgE
```

```
order by StartAgE

---2166 patietns don't seem to have a startage, there are also a lot of  patients under 16
---Lets look at the patients who do not have a start age
select *
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where STARTAGE is null
---2166 rows
---quite a lot of these have a drop record
select *
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where STARTAGE is null  and droprecord is null
---only 2559 rows here
select Patient_ID, StartAge
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where Patient_ID in (select distinct Patient_ID
                                        from Cancer_research.dbo.FE_20131018_AllColonoscopies
                                        where STARTAGE is null  and droprecord is null )
---no additional age groups here...
---Lets have a look to see whether there are any ages for these patietns on HES

select t1.Patient_ID, t1.StartAge, ADMIAGE, ACTIVAGE, ENDAGE
from HED.dbo.HES_Episodes t1 inner join (select distinct Patient_ID, spell_ID
                                                                                from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                                where
STARTAGE is null  and droprecord is null )t2
                                        on t1.Patient_ID = t2.Patient_ID
                                        and t1.SPELL_ID = t2.SPELL_ID
----lets update these now...  there are some additional ages here
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set STARTAGE = Admiage
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select t1.Patient_ID,
t1.StartAge, ADMIAGE, ACTIVAGE, ENDAGE

                                        from HED.dbo.HES_Episodes t1 inner join (select distinct
Patient_ID, spell_ID


                                from Cancer_research.dbo.FE_20131018_AllColonoscopies


                                where STARTAGE is null  and droprecord is null )t2

                                                                                on t1.Patient_ID =
t2.Patient_ID

                                                                                and t1.SPELL_ID =
t2.SPELL_ID) t2
                                                                                on t1.Patient_ID =
t2.Patient_ID
where t1.STARTAGE is null
----(868 row(s) affected)

---have a look in the first hes episodes tab;e
select t1.Patient_ID, t1.StartAge, ADMIAGE, ACTIVAGE, ENDAGE
from HED.dbo.HES_Episodes_Part1 t1 inner join (select distinct Patient_ID, spell_ID
                                                                                from
Cancer_research.dbo.FE_20131018_AllColonoscopies
                                                                                where
STARTAGE is null  and droprecord is null )t2
                                        on t1.Patient_ID = t2.Patient_ID
                                        and t1.SPELL_ID = t2.SPELL_ID ---nope nothing
---lets remove the rest of the records with a null start age

update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord = 1
where STARTAGE is null or STARTAGE = 999
----(1387 row(s) affected)
---after speaking to Ravi he suggested that the best hting to do would be to remove those patients
under 18
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord = 1
where Patient_ID in (select distinct Patient_ID
```

```
                                    from  Cancer_research.dbo.FE_20131018_AllColonoscopies
                                    where STARTAGE < 18 or STARTAGE > 200)---46008 rows
---- lets have a look at the recors where there is n't atrecord for colonoscopy date
select *
from Cancer_research.dbo.FE_20131018_AllColonoscopies
where colonoscopy_date is null and droprecord is null
----4842
---some of these are obviously day case appointments
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set colonoscopy_date = DISDATE
where EPISTART = DISDATE and colonoscopy_date is null
---3984
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set colonoscopy_date = EPISTART
where colonoscopy_date is null
---1134
 ---there are soem colonoscopies which occur outside the date range
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord  = 1
where colonoscopy_date <= '20010331'
---     (298 row(s) affected)

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set droprecord  = 1
where colonoscopy_date > '20120331'
----    (324 row(s) affected)
---Need to add ethnicity into this tabel
alter table Cancer_research.dbo.FE_20131018_AllColonoscopies
add  ethnic_group varchar(50),
ethnos varchar(5);


---Need to add ethnicity to these tables
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set ethnos = t2.ethnos
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join HED.dbo.HES_EPISODES t2
                                                            on t1.Patient_ID = t2.Patient_ID
                                                            and t1.SPELL_ID = t2.SPELL_ID
                                                            and t1.EPIORDER = t2.EPIORDER

-----(1754899 row(s) affected)

update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set ethnos = t2.ethnos
from  Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join HED.dbo.HES_EPISODES_part1 t2
                                                            on t1.Patient_ID = t2.Patient_ID
                                                            and t1.SPELL_ID = t2.SPELL_ID
                                                            and t1.EPIORDER = t2.EPIORDER
---(1178389 row(s) affected)

---lets look at these ethnicities
select ethnos, COUNT(*)
from  Cancer_research.dbo.FE_20131018_AllColonoscopies
where droprecord is null
group by ethnos
order by ethnos
---there are quite  a  lot here which are invalid
---lets have a look whether there are any better ethnicities and any patietns who only have one
ethnicity recorded


select  t1.patient_ID, COUNT(distinct t2.ethnos)
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1  inner join (select Patient_ID, Ethnos,
admidate

                                                        from HED.dbo.HES_EPISODES_Part1

                                                        where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')

                                                        union all

                                                        select Patient_ID, Ethnos , admidate
```

```
                                                    from HED.dbo.HES_EPISODES

                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2

        on t1.Patient_ID = t2.Patient_ID
where t1.ethnos  in ('0','1','2','3','4','5','6','7','8','9','x','y','z')
group by t1.Patient_ID
having COUNT(distinct t2.ethnos) =1
----311477
---updating these patietns
update Cancer_research.dbo.FE_20131018_AllColonoscopies
set ethnos = t2.ETHNOS
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select Patient_ID, Ethnos,
admidate

                                                    from HED.dbo.HES_EPISODES_Part1

                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')

                                                            and Patient_ID in (select
t1.patient_ID

        from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select Patient_ID,
Ethnos, admidate


                                                    from HED.dbo.HES_EPISODES_Part1


                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')


                                                    union all


                                                    select Patient_ID, Ethnos , admidate


                                                    from HED.dbo.HES_EPISODES


                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2


                on t1.Patient_ID = t2.Patient_ID

        where t1.ethnos  in ('0','1','2','3','4','5','6','7','8','9','x','y','z')

        group by t1.Patient_ID

        having COUNT(distinct t2.ethnos) =1 )

                                                    union all

                                                    select Patient_ID, Ethnos , admidate

                                                    from HED.dbo.HES_EPISODES
```

```
                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')

                                                     and Patient_ID in (select
t1.patient_ID

        from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select Patient_ID,
Ethnos, admidate

                                                    from HED.dbo.HES_EPISODES_Part1

                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')

                                                    union all

                                                    select Patient_ID, Ethnos , admidate

                                                    from HED.dbo.HES_EPISODES

                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2

                on t1.Patient_ID = t2.Patient_ID

        where t1.ethnos  in ('0','1','2','3','4','5','6','7','8','9','x','y','z')

        group by t1.Patient_ID

        having COUNT(distinct t2.ethnos) =1 )) t2

                on t1.Patient_ID = t2.PAtient_ID
---(471885 row(s) affected)
---having a look at the patients now, who have a proper ethnicity, but more than one recorded
select t1.Patient_ID, t2.ethnos, COUNT(distinct t2.ADMIDATE) admissions
into #ethnicmastectomy
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select Patient_ID, Ethnos,
admidate

                                                    from HED.dbo.HES_EPISODES_Part1

                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')

                                                    union all

                                                    select Patient_ID, Ethnos , admidate

                                                    from HED.dbo.HES_EPISODES

                                                    where ETHNOS not in
('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2

        on t1.Patient_ID = t2.Patient_ID
where  t1.patient_ID in (select  t1.patient_ID
```

```
                                                from Cancer_research.dbo.FE_20131018_AllColonoscopies
t1 inner join (select Patient_ID, Ethnos, admidate

        from HED.dbo.HES_EPISODES_Part1

        where ETHNOS not in ('0','1','2','3','4','5','6','7','8','9','x','y','z')

        union all

        select Patient_ID, Ethnos , admidate

        from HED.dbo.HES_EPISODES

        where ETHNOS not in ('0','1','2','3','4','5','6','7','8','9','x','y','z')) t2

                                                        on t1.Patient_ID = t2.Patient_ID
                                                where t1.ethnos  in
('0','1','2','3','4','5','6','7','8','9','x','y','z')
                                                group by t1.Patient_ID
                                                having COUNT(distinct t2.ethnos) >1 )
group by t1.Patient_ID, t2.ethnos
-----(50175 row(s) affected)

select distinct PAtient_ID
from #ethnicmastectomy;

----24060
select t1.*
from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions) maxadmi
                                                                                        from
#ethnicmastectomy

        group by Patient_ID) t2
                                                                on t1.Patient_ID = t2.Patient_ID
                                                                and t1.admissions = t2.maxadmi
---     28337

select Patient_ID, COUNT(distinct ethnos) ethnicities
into #ethnicUniquePatient
from (select t1.*
                from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions) maxadmi

                from  #ethnicmastectomy

                group by Patient_ID) t2
                                                                        on t1.Patient_ID =
t2.Patient_ID
                                                                        and t1.admissions =
t2.maxadmi) a
group by Patient_ID
having COUNT(distinct ethnos) = 1
----19915

update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set ethnos = t2.ethnos
from   Cancer_research.dbo.FE_20131018_AllColonoscopies t1    inner join  (select t1.*

                        from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions)
maxadmi

                                                from  #ethnicmastectomy

                                                group by Patient_ID) t2

                on t1.Patient_ID = t2.Patient_ID
```

```
                and t1.admissions = t2.maxadmi
                        where t1.Patient_ID in (select distinct Patient_ID
                                                from #ethnicUniquePatient) ) t2
                                                                        on
t1.PAtient_ID = t2.Patient_ID
where t1.ethnos in ('0','1','2','3','4','5','6','7','8','9','x','y','z')

---(22619 row(s) affected)

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set ethnos = t2.ethnos
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join (select t1.* , t4.lastadmission
                from #ethnicmastectomy t1  inner join (select Patient_ID, MAX(admissions) maxadmi
                                        from  #ethnicmastectomy
                                        group by Patient_ID) t2
        on t1.Patient_ID = t2.Patient_ID
        and t1.admissions = t2.maxadmi
                                                                inner join (select
Patient_ID, Ethnos, MAX(admidate)lastadmission
        from (select Patient_ID, admidate , Ethnos
                from HED.dbo.HES_Episodes
                union all
                select Patient_ID, admidate, Ethnos
                from HED.dbo.HES_EPISODES_Part1) a
        group by Patient_ID, Ethnos) t3
                                                                                on
t1.Patient_ID = t3.Patient_ID
                                                                                and
t1.ETHNOS = t3.ETHNOS
                                                                inner join (select
Patient_ID, MAX(admidate)lastadmission
        from (select Patient_ID, admidate , Ethnos
                from HED.dbo.HES_Episodes
                union all
                select Patient_ID, admidate, Ethnos
```

```
                from HED.dbo.HES_EPISODES_Part1) a

        group by Patient_ID) t4
                                                                                    on
t1.Patient_ID = t4.Patient_ID
                                                                                    and
t3.lastadmission = t4.lastadmission
                where t1.Patient_ID not in (select distinct Patient_ID
                                                                                        from
#ethnicUniquePatient) ) t2
                                                                                  on
t1.Patient_ID = t2.Patient_ID
where t1.ethnos in ('0','1','2','3','4','5','6','7','8','9','x','y','z')
        ----    (3827 row(s) affected)



---lets check out the ethnicites again
select ethnos, COUNT(*)
from Cancer_research.dbo.FE_20131018_AllColonoscopies
group by ethnos
order by ethnos
---this is a huge improvement on the previous data - still some old codes though...

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set ethnic_group = case     when ethnos in ('A','B','C','0') then 'White'
                                        when ethnos in ('D','E','F','G') then 'Mixed'
                                        when ethnos in ('H','J','K','L','4','5','6') then
'Asian or Asian British'
                                        when ethnos in ('M','N','P','1','2','3') then 'Black
or Black British'
                                        when ethnos in ('R','7') then 'Chinese'
                                        when ethnos in ('S','8') then 'Other Ethnic Groups'
                                        when ethnos in ('Z','X','9') then 'Unknown' end

        ---(2933288 row(s) affected)
alter table Cancer_research.dbo.FE_20131018_AllColonoscopies
add charlson int;

update Cancer_research.dbo.FE_20131018_AllColonoscopies
set charlson = t2.DOMI_0214_CHARLSON_NWT
from Cancer_research.dbo.FE_20131018_AllColonoscopies t1 inner join HED.dbo.HES_SPELLS t2
                                                        on t1.Patient_ID = t2.PATIENT_ID
                                                        and t1.SPELL_ID = t2.SPELL_ID


----(2931930 row(s) affected)

alter table Cancer_research.dbo.FE_20131018_AllColonoscopies
add type_colonoscopy varchar(25);

set concat_null_yields_null off
update  Cancer_research.dbo.FE_20131018_AllColonoscopies
set type_colonoscopy = case when
OPERTN_01+OPERTN_02+OPERTN_03+OPERTN_04+OPERTN_05+OPERTN_06+OPERTN_07+opertn_08+OPERTN_09+OPERTN_10+OP
ERTN_11+OPERTN_12 like '%H20[16]%' then 'Polypectomy'
                                                        when
OPERTN_01+OPERTN_02+OPERTN_03+OPERTN_04+OPERTN_05+OPERTN_06+OPERTN_07+opertn_08+OPERTN_09+OPERTN_10+OP
ERTN_11+OPERTN_12 like '%H221%' then 'Biopsy'    else 'Other' end

select *
from Cancer_research.dbo.FE_20131018_AllColonoscopies
```